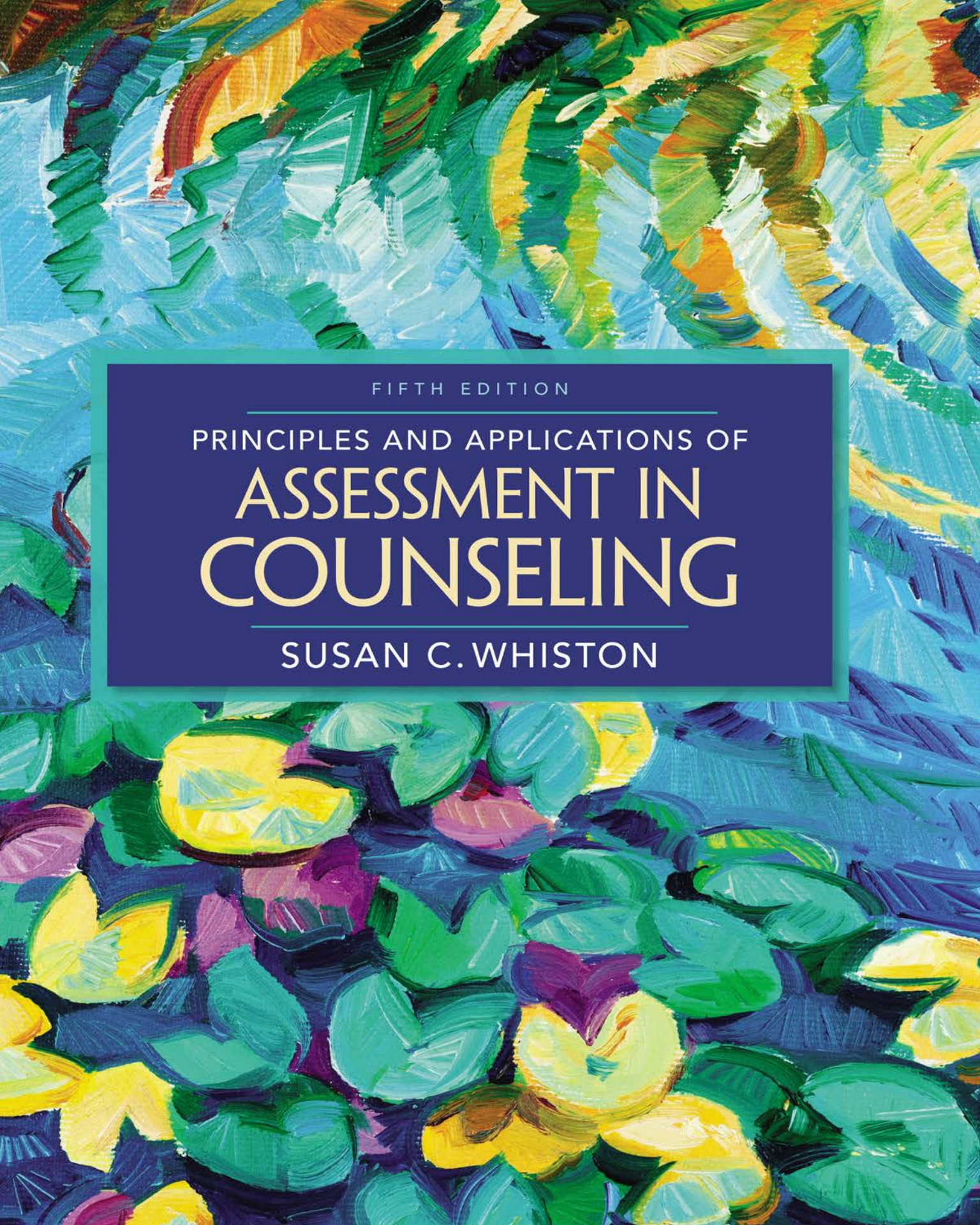

FIFTH EDITION
PRINCIPLES AND APPLICATIONS OF
ASSESSMENT IN
COUNSELING
SUSAN C. WHISTON

Want to turn C's into A's? Obviously, right?

But the right way to go about it isn't always so obvious. Go digital to get the grades. MindTap's customizable study tools and eTextbook give you everything you need all in one place.

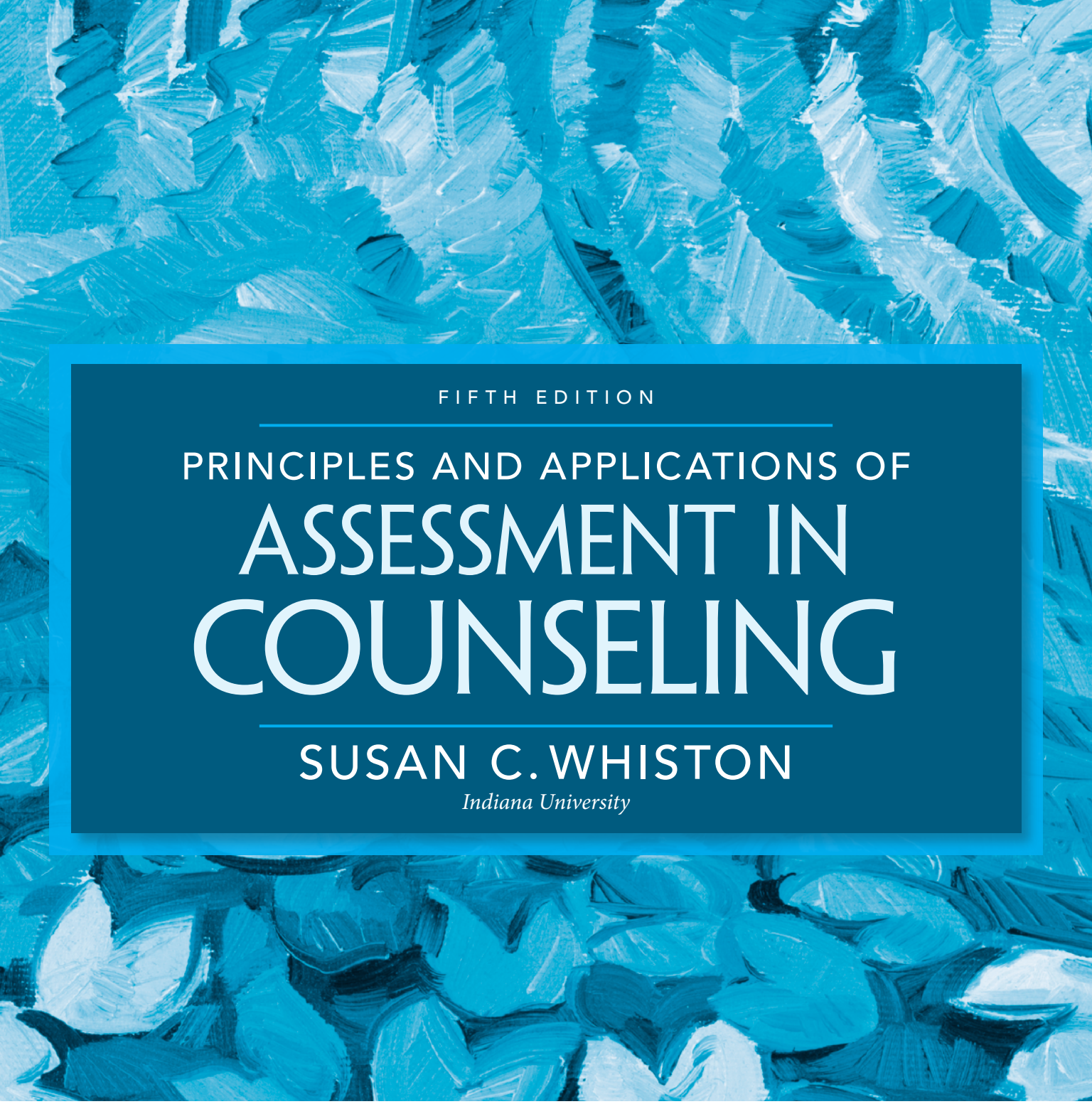
Engage with your course content, enjoy the flexibility of studying anytime and anywhere, stay connected to assignment due dates and instructor notifications with the MindTap Mobile app...

*and most of all...***EARN BETTER GRADES.**

TO GET STARTED VISIT
WWW.CENGAGE.COM/STUDENTS/MINDTAP

CENGAGE Learning® | MindTap®

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.



FIFTH EDITION

PRINCIPLES AND APPLICATIONS OF ASSESSMENT IN COUNSELING

SUSAN C. WHISTON
Indiana University

CENGAGE Learning®

Australia • Brazil • Mexico • Singapore • United Kingdom • United States

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit www.cengage.com/highered to search by ISBN#, author, title, or keyword for materials in your areas of interest.

Important Notice: Media content referenced within the product description or the product text may not be available in the eBook version.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CENGAGE Learning®

***Principles and Applications of Assessment in Counseling,* Fifth Edition**
Susan C. Whiston

Product Director: Jon-David Hague

Product Manager: Julie A Martinez

Content Developer: Elizabeth Momb

Managing Content Developer: Mary Noel

Product Assistant: Stephen Lagos

Marketing Manager: Margaux Cameron

Art and Cover Direction, Production Management, and Composition: Lumina Datamatics, Inc.

Manufacturing Planner: Judy Inouye

Cover Image: © Boyan Dimitrov/ Shutterstock

Unless otherwise noted, all items
© Cengage Learning

© 2017, 2013 Cengage Learning

WCN: 02-200-203

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced or distributed in any form or by any means, except as permitted by U.S. copyright law, without the prior written permission of the copyright owner.

For product information and technology assistance, contact us at
Cengage Learning Customer & Sales Support, 1-800-354-9706.

For permission to use material from this text or product, submit all requests online at **www.cengage.com/permissions.**
Further permissions questions can be e-mailed to
permissionrequest@cengage.com.

Library of Congress Control Number: 2015948077

Student Edition:
ISBN: 978-1-305-27148-7

Loose-leaf Edition:
ISBN: 978-1-305-86636-2

Cengage Learning
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with employees residing in nearly 40 different countries and sales in more than 125 countries around the world. Find your local representative at **www.cengage.com.**

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions, visit **www.cengage.com.**

Purchase any of our products at your local college store or at our preferred online store **www.cengagebrain.com.**

Printed in the United States of America
Print Number: 01 Print Year: 2016

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

To Jen, Michael, and Matt

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Brief Contents

Preface ix

CHAPTER 1 **Assessment in Counseling** 1

CHAPTER 2 **Basic Assessment Principles** 18

CHAPTER 3 **Reliability** 37

CHAPTER 4 **Validity and Item Analysis** 56

CHAPTER 5 **Ethical and Legal Issues in Assessment** 78

CHAPTER 6 **Issues Related to Assessment with Diverse Populations** 96

CHAPTER 7 **Selecting, Administering, Scoring, and Communicating Assessment Results** 119

CHAPTER 8 **Initial Assessment in Counseling** 140

CHAPTER 9 **Intelligence and General Ability Testing** 168

CHAPTER 10 **Assessing Achievement and Aptitude: Applications for Counseling** 193

CHAPTER 11 **Assessment in Career Counseling** 214

CHAPTER 12 **Appraisal of Personality** 237

CHAPTER 13 **Behavioral Assessment** 267

CHAPTER 14 **Assessment in Marriage and Family Counseling** 281

CHAPTER 15 **Diagnosis** 300

CHAPTER 16 **Monitoring and Evaluating Counseling** 328

Appendix A 2014 ACA Code of Ethics Section E 342

Appendix B Table of Areas Under the Normal Curve 347

Appendix C Standards for Multicultural Assessment (4th Edition) 349

Glossary 355

References 362

Index 379

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Contents

Preface ix

CHAPTER 1
Assessment in Counseling 1
What Is Assessment? 2
Do Counselors Need to Know About Assessment? 3
Assessment Is Integral to Counseling 4
Assessment Can Be Therapeutic 7
What Do Counselors Need to Know About Assessment? 8
Types of Assessment Tools 9
History 11
Summary 17

CHAPTER 2
Basic Assessment Principles 18
Measurement Scales 19
Norm-Referenced vs. Criterion-Referenced Instruments 20
Interpreting Norm-Referenced Instruments 21
Measures of Central Tendency 23
Measures of Variability 25
Normal Distributions 27
Standard Scores 30
Stanines 32
Evaluating the Norming Group 35
Summary 36

CHAPTER 3
Reliability 37
Classical Test Theory 38
Correlation Coefficients 39
Types of Reliability Coefficients 43
Evaluating Reliability Coefficients 47
Standard Error of Measurement 48
Standard Error of Difference 51
Generalizability Theory 52
Item Response Theory 53
Summary 54

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 4
Validity and Item Analysis 56
Validity 56
Item Analysis 71
Summary 77

CHAPTER 5
Ethical and Legal Issues in Assessment 78
Ethics 79
ACA Code of Ethics 80
Client Rights and Responsibilities 85
Counselors Rights and Responsibilities 86
Legal Issues in Assessment 87
Summary 95

CHAPTER 6
Issues Related to Assessment with Diverse Populations 96
Culture and Assessment 98
Assessment of Individuals with Disabilities 109
Summary 117

CHAPTER 7
Selecting, Administering, Scoring, and Communicating Assessment Results 119
Selection of Assessment Instruments or Strategies 120
Administering Assessment Instruments 126
Scoring 126
Communicating Results 129
Summary 138

CHAPTER 8
Initial Assessment in Counseling 140
Initial Intake Interview 141
Defining the Client's Problem 144
Interviewing Skills and Techniques 145
Interviewing Children 146
Structured and Unstructured Initial Interviews 147
Strengths and Limitations of Interviewing 149
Other Strategies That Can Be Used in Initial Assessment 150
Assessing Specific Problems in the Initial Sessions 153
Mental Status Examination 164
Summary 166

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 9
Intelligence and General Ability Testing 168
Models of Intelligence 169
Individual Intelligence Testing 174
Group Intelligence Testing 184
Issues in Intelligence Testing 186
Summary 191

CHAPTER 10
Assessing Achievement and Aptitude: Applications for Counseling. 193
Assessment of Achievement 194
Aptitude Assessment 201
Test Preparation and Performance 211
Summary 212

CHAPTER 11
Assessment in Career Counseling. 214
Assessing Individual Differences 216
Career Process Measures 227
Qualitative Career Assessment 229
Issues and Trends in Career Assessment 231
Summary 235

CHAPTER 12
Appraisal of Personality 237
Informal Assessment Techniques 239
Structured Personality Inventories 241
Projective Techniques 260
Self-Concept Measures 265
Summary 266

CHAPTER 13
Behavioral Assessment 267
Observational Recording Methods 270
Self-Monitoring 275
Reactivity 276
Formal Behavioral Assessments 277
Error in Behavioral Assessment 278
Summary 279

CHAPTER 14
Assessment in Marriage and Family Counseling 281
Difficulties with Assessment of Couples and Families 283
Family Interviews 285

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Mapping Techniques 286
Use of Instruments Designed for Individuals 291
Formalized Assessment Instruments Specific to Marriage and Family 292
Summary 298

CHAPTER 15
Diagnosis **300**
DSM-5's Definition of a Mental Disorder 302
Organizational Structure 303
Cultural Issues 304
Nonaxial System 304
Other Specified and Unspecified Disorders 305
Subtypes and Specifiers 305
DSM-5 Emerging Measures 306
Diagnostic Criteria 307
Common Diagnoses 309
Implications for Counselors 324
Instruments Designed to Assist in Diagnosis 324
Summary 327

CHAPTER 16
Monitoring and Evaluating Counseling **328**
Treatment Monitoring 329
Evaluation in Counseling 333
Summary 340

Appendix A 2014 ACA Code of Ethics Section E 342
Appendix B Table of Areas Under the Normal Curve 347
Appendix C Standards for Multicultural Assessment (4th Edition) 349
Glossary 355
References 362
Index 379

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Preface

Consistent with previous editions of *Principles and Applications of Assessment in Counseling*, this edition is based on the premise that counselors functioning in almost all clinical or school settings must have a sound understanding of assessment. I will argue that assessment is an integral part of the counseling process and that all practitioners need refined assessment skills. In my opinion, selecting a treatment or determining the appropriate counseling theory should be based on a sound and comprehensive assessment of the client. Sound assessments are not used solely at the beginning of the counseling process; they are also used to monitor treatment and evaluate therapeutic services at the conclusion of counseling. Sound assessments occur when a counselor gathers accurate information and interprets and uses that information in an appropriate manner. The importance of sound assessment procedures applies to both formal and informal assessment techniques that are used throughout the counseling process. The goals of this book are, first, to provide a foundation in assessment so clinicians can select and use sound assessment techniques and, second, to expose counselors to a variety of assessment areas and issues.

New to This Edition

The field of assessment in counseling is continually changing and evolving; hence, this edition was initiated to reflect many of those changes. A significant influence on the field of assessment is the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014), and a new edition of this resource was published since the last edition of this book. This resource, which is often referred to as the *Standards*, is frequently considered the guidebook in the area of assessment, and its content had a significant influence on many of the changes in my book. In particular, new information is infused into the first four chapters of this book. The infusion of new information, however, is consistent with the other editions of this book in terms of keeping the coverage of basic measurement concepts fairly simple and understandable.

One of the major changes to this edition of the book is the organization of the chapters. Consistent with the older versions of the book, the first four chapters address basic measurement concepts that provide a foundation for the remainder of the book. New to this edition, the fifth chapter addresses ethics and legal issues in assessment. A few of my colleagues have suggested moving this chapter earlier in the book, and I have followed their advice in this fifth edition. This chapter on ethics and legal issues includes information from the *2014 ACA Code of Ethics* (American Counseling Association, 2014). With this edition of the book, the chapter on ethics and legal issues is followed by a chapter related to assessment issues in working with diverse populations. Given the changing demographics in the United States and the need for assessment in counseling internationally, this is an important chapter.

New to this edition of the book is an expansion of the chapter related to initial assessment. This chapter was primarily rewritten to correspond to the 2016 Council for Accreditation of Counseling and Related Educational Programs (CACREP) Standards. Another

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

change concerns the chapter that addresses the assessment of intelligence or general ability, which now includes information on the recently released *Wechsler Intelligence Scale for Children—Fifth Edition* (WISC-V). The chapter that covers assessing achievement and aptitude assessment focuses on new developments in the area of achievement testing and includes information on the new SAT. I have included an entirely new chapter to this edition, which addresses the topic of behavioral assessment. This addition was also in response to the 2016 CACREP Standards that require students to be competent in systematic behavioral observations.

Diagnostic skills are important for all counselors, and Chapter 15 contains an overview of the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition* (DSM-5). The DSM-5 is the most widely used diagnostic system of mental disorders in the United States, and this chapter introduces the reader to the DSM-5 and to the disorders most commonly diagnosed by counselors. Also new to this edition of the book is a chapter devoted exclusively to the monitoring and evaluation of counseling. There is compelling research that counselors who receive consistent feedback on client progress tend to have better outcomes as compared to counselors who do not receive feedback; therefore, monitoring client progress is important. In addition, in these days of accountability counselors need to be able to evaluate counseling services and assessment plays a critical role in the evaluation process, which is also addressed in this chapter.

Overview of the Book

In the first four chapters of this book, I focus on the underlying measurement principles of any type of psychological assessment. This section is designed so that clinicians have a foundation in basic measurement concepts and the fundamentals of reliability and validity. The goal is to provide fundamental information that can be used to evaluate many types of formal and informal assessment techniques. Although the intent of this book is not to turn readers into statisticians, there are a few formulas. Readers, however, are encouraged to focus more on concepts than on mathematical calculations. For example, in most instrument manuals, there will be standard deviation(s), and counselors will not need to calculate a standard deviation. They will, however, need to understand how the standard deviation is calculated and what it means. Chapter 1 introduces assessment terms and briefly describes some of the historical influences on assessment practices and procedures. To use many assessment instruments properly, counselors need an understanding of some basic statistics that *describe* individuals' scores. Therefore, Chapter 2 includes a discussion of the two basic types of assessment instruments (i.e., norm-referenced and criterion-referenced) and some fundamental information on measurement and scoring.

Furthermore, counselors also need to understand the psychometric qualities of assessment instruments. In general, the psychometric qualities considered by experts in assessment to be of primary importance are *reliability* and *validity*. In counseling, these psychometric qualities are important as they help the clinician decide if and how an assessment technique should be used. Chapters 3 and 4 cover these two topics as well as how they are important in both evaluating instruments and determining the meaning of the assessment results.

Because the professional responsibilities associated with assessments are substantial, any professional counselor must have comprehensive knowledge of ethical practices and professional standards related to counseling assessment. Chapter 5 of the book addresses the importance of ethical and legal practices in assessment. In particular, there is a focus on the *2014 ACA Code of Ethics* that should guide every counselor's practice. This chapter also

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

includes an overview of legislation and litigation cases that are pertinent to assessment. Competent assessment also requires knowledge about the influence of background and culture, particularly as it relates to race and ethnicity and other issues of diversity. Chapter 6 provides an overview of topics relevant to the assessment of diverse clients. Other responsibilities of counselors in the assessment process are the selection of an instrument, administering the instrument, scoring it, and communicating the results to a client. These activities are covered in Chapter 7.

Chapter 8, which discusses initial assessment in counseling, focuses on topics related to the initial interview and to assessment methods that can be used during the first session with a client. This chapter addresses topics such as how to assess for possible suicidal ideation. Chapters 9 and 10 address the assessment of human cognitive abilities: Chapter 9, the measure of intelligence or general ability; and Chapter 10, achievement and aptitude assessment. Counselors are expected to be knowledgeable in these areas and to be able to interpret the results from instruments that measure them. Our society is changing rapidly, and these changes influence work and employment, resulting in many individuals seeking career counseling. Because assessment is often a part of career counseling, this area is addressed in Chapter 11. Chapter 12 discusses personality assessment and many of the methods used to assess personality. Chapter 13 is a new chapter, which encompasses the topic of behavioral assessment and particularly how to conduct systematic behavioral observations. There is also a chapter, Chapter 14, which focuses on assessing couples and families.

An important primary step in the counseling process is often assessment and diagnosis. Accurate diagnosis of client issues and problems is central to the counseling process—if a counselor misdiagnoses or does not properly assess the client's issues in the initial stages, then the chances of providing effective counseling services are diminished. In many mental health settings, counselors must have refined diagnostic skills. Therefore, Chapter 15 is designed as an introduction to the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition* (DSM-5), a commonly used diagnostic system. The book concludes with a discussion of the importance of monitoring client progress and how assessments are used in this monitoring process. Chapter 16 concludes with a summary of how to use assessments in the evaluation of the effectiveness of the counseling.

Acknowledgments

I wrote the first edition of this book while on sabbatical at the University of Nevada, Las Vegas. Although many people from Las Vegas contributed to the first edition of this book, one of my colleagues, Fred Kirschner, deserves a special note of gratitude. Fred's guidance during the first edition of the book continued to influence my writing of this edition.

In all editions of this book, I have attempted to write from a student perspective. I need to thank my students at Indiana University whom I continue to learn from and who influenced this edition of the book. The students at Indiana University are an outstanding group, and I am lucky to have the opportunity to work with these exemplary emerging professionals. I also want to thank Julie Martinez and Elizabeth Momb who were extremely supportive of a fifth edition of this book. In addition, a number of reviewers contributed valuable input and suggestions; among them were Nancy Leech, University of Colorado Denver; Mary Deacon, Liberty University; Regina Kakhnovets, Austin Peay State University; Joshua Gold, University of South Carolina; Edward Wahesh, Villanova University; Edward Keane, Housatonic Community College; and Stephanie Bailey, Loyola Marymount University. Their assistance is much appreciated.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

This edition of the book is dedicated to my children, Jen, Michael, and Matt, who are all maturing into wonderful people in spite of their mother. I have continued to assess them since they were babies and I always have been happy with the results. I would also like to thank Kevin, who served as a resource on content and who was extremely supportive during the writing of this edition of the book.

Susan C. Whiston
Indiana University

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 1

Assessment in Counseling

Learning Objectives

- LO1 Define the term *assessment*
- LO2 Debate whether assessment is an integral part of counseling
- LO3 Understand the Council for the Accreditation of Counseling and Related Educational Programs' (CACREP) standards for assessment
- LO4 Distinguish among the different types of assessment tools
- LO5 Comprehend major historical influences on today's practices in assessment

By their very nature, humans are complex, and the more pieces of the human puzzle clinicians can access, the more likely it is that they will have a more complete picture. As with a jigsaw puzzle, if counselors have only a few pieces of the puzzle, they will have a hard time determining what the whole picture is. Formal and informal assessment techniques can assist counselors in gathering "puzzle pieces," thus facilitating a more comprehensive view of clients. Counselors with numerous methods for accessing client information have more material to use in a therapeutic manner than do those who have limited assessment skills.

As a counselor, you will regularly assess clients. This assessment may address what the client's issues are, the magnitude of his or her problems, the client's strengths and level of motivation, or whether counseling can be beneficial. Assessing people is probably not entirely new to you. Chances are you have been doing assessment on an informal basis for many years. When you walk into a class for the first time, you may begin to casually assess your fellow classmates. You may have also assessed personalities or interests of people at a social gathering. Assessing other individuals is part of everyday life; it is also part of the counseling process. In the 1930s and 1940s, many people viewed the terms *counseling* and *assessment* as being synonymous and most of the public thought counseling always involved the use of tests (Hood & Johnson, 2007). The majority of clinicians currently do not see counseling and assessment as synonymous; however, client assessment does continue to play a crucial role in the counseling process.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

LO1 What Is Assessment?

Before further pursuing the topic of assessment in counseling, it is important to discuss the precise meaning of assessment. A term closely associated with assessment is **psychological test**, which Anastasi and Urbina (1997) defined as an objective and standardized measure of a sample of behavior. Cronbach's (1990) definition is similar, with a test being a systematic procedure for observing behavior and describing it with the aid of numerical scales or fixed categories. As these two similar definitions suggest, many of the definitions of assessment have some commonalities. They all discuss getting a measure or using some type of measurement. In assessment, counselors often want an indication of quantity (e.g., How depressed is the client? Are the test scores high enough to get into an Ivy League school?). In simple terms, in assessment we are often attempting to determine whether there is a lot of "something" or just a little.

The issue of quantity leads to the next topic: the something that is being measured. In counseling, practitioners are often interested in human constructs such as emotions, career interests, personality factors, abilities, and aptitudes. These *constructs*, however, are difficult to measure directly. For example, individuals cannot give a pint of emotions in the same way they can a pint of blood. Humans, for the most part, indicate their emotions by their behaviors, their statements, or even the answers they give on a questionnaire. It is important to remember that speaking and responding to a questionnaire are behaviors. Even when people take a test or an assessment, their answers are just a sample of behaviors in that area. Thus, for most of the areas in which counselors want to gather information, all that they can truly gather are *samples of behavior* (although in very rare instances, a clinician may use a physiological measure, such as pulse rate as a measure of anxiety). In addition, when assessing clients, counselors observe a sample of behavior and then *infer* certain meanings from those behaviors. Hence, when first beginning to observe or assess a client, counselors should consider their observations as a sample of behavior and then reflect on two important questions: First, is the sample of behavior indicative of how the person usually behaves, and second, are the inferences being made about this sample of behavior correct?

If the intent is to obtain a sample of behaviors and make some inferences or clinical decisions, then it makes sense for counselors to be careful about the manner in which they obtain the behavior samples. Being careful during the assessment process is particularly important when the intent of the assessment process is to compare individuals and make decisions (e.g., to determine who gets a scholarship or whether a client has a substance abuse problem). This care is related to the third common point in the definitions of assessment—an objective or systematic measure of behavior. For example, if a counselor wants to assess clients' level of motivation, then she would need to gather samples of behaviors that reflect each client's level of motivation. The counselor might decide that a direct approach is good, so she would ask each client about the level of motivation. To the first client, she might ask, "Are you motivated to try some different things in counseling?" To the second client, she might ask, "Are you motivated?" With the third client, she might say, "You do not seem very motivated. Are you?" This counselor's way of gathering samples of behavior will probably affect the behaviors of the clients, so their answers may be quite different depending on the way she asks the question. This illustrates how important it is that, in assessment, the manner of gathering behavior samples be objective, standardized, and systematic.

Although this book's focus is on assessment and appraisal in counseling, it does adhere to the traditional definition of psychological assessment. To be fair to all clients, assessments need to be systematic and objective. With assessment in counseling, clinicians are, in essence, gathering samples of client behaviors and making inferences based on those

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

behaviors. Therefore, when considering assessment strategies, clinicians should focus on the methods or procedures used for gathering the samples of behavior. For example, a counselor interviewing a client should consider what behaviors are being sampled and whether these behaviors are typical of that client. The counselor also needs to consider the inferences he makes about the client and the evidence (or validity) of those inferences.

The final aspect of assessment is measurement. When assessing a client, a counselor is attempting to measure some aspect of the client. Even if the question is whether or not the client has some particular attribute, the assessment involves measurement. For instance, determining whether a client is suicidal involves measuring the degree to which suicidal indicators are present. One of the challenges in assessment is finding methods for obtaining meaningful measures of many psychological constructs. For example, we cannot say a client has one liter of anxiety.

A distinction is sometimes made among the words **assessment**, **appraisal**, and **testing**. The *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), which is one of the primary authoritative sources related to assessment, distinguishes between the terms *test* and *assessment*. It defines assessment as "a broader term than test, commonly referring to a process that integrates test information with information from other sources (e.g., information from other tests, inventories, and interviews; or the individual's social educational, employment, health or psychological history)" (p. 2). In this book, the terms *assessment* and *appraisal* are used interchangeably, based on the opinion that they both include the use of formal and informal techniques, not just standardized tests. Assessment and appraisal are not just formal psychological evaluations; in this book, they are defined as procedures for gathering client information that is then used to facilitate clinical decisions and provide information to clients. A distinction, however, does need to be made between **tests** and **instruments**. The term *test* is often reserved for an individual instrument in which the focus is on evaluation, such as a test that is graded in a class. Many instruments that counselors use, such as scales, checklists, and inventories, are designed to provide information and are not evaluative. In this book, the term *instrument* includes tests, scales, checklists, and inventories. As you read this book, there may be terms with which you are unfamiliar; in such a case, the glossary at the end of the book may be helpful.

Do Counselors Need to Know About Assessment?

Counselors should know about assessment because the American Counseling Association's (ACA) (2014) *Code of Ethics* devoted an entire section (Section E) to evaluation, assessment, and interpretation. Furthermore, the Council for Accreditation of Counseling and Related Educational Programs (CACREP, 2009) stipulated that eight common core areas are required of all students in accredited counseling programs, one of which is assessment. Furthermore, assessment is one of the major topics on the National Counselor Examination (National Board for Certified Counselors, 2014). Hence, there are professional expectations that all counselors have training and knowledge in assessment and in the appropriate use of assessment instruments and techniques.

Numerous research studies indicate that counselors in a variety of settings (e.g., schools, community agencies, mental health facilities) view formal assessment strategies as a significant aspect of their work. Ekstrom, Elmore, Schafer, Trotter, and Webster (2004) surveyed school counselors. Ninety-one percent indicated they interpreted scores from tests/assessments and used the information in counseling. These researchers also found that

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

81% communicated and interpreted test/assessment information to parents and 80% had communicated and interpreted test/assessment information to teachers, school administrators, and other professionals. With the advent of school reform, school counselors are even more involved in the administration and interpretation of tests. With a sample of national certified counselors, Peterson, Lomas, Neukrug, and Bonner (2014) found that the most commonly used instruments were the Beck Depression Inventory, the Myers–Briggs Type Indicators, the Strong Interest Inventory, the ACT, the SAT/PSAT, and the Self-Directed Search.

Refined assessment skills can often assist counselors in providing treatment more quickly and efficiently. Meyer et al. (2001) compared the empirical evidence concerning psychological assessments and medical testing and found that psychological assessments are comparable to medical tests in terms of accuracy. In addition, the public expects that counselors understand assessment and are able to interpret assessment results. Both mental health and school counselors work with other professionals in multidisciplinary teams where other team members (e.g., psychologists) will quickly lose respect for a counselor who has limited knowledge of psychometric principles and common assessment instruments. Moreover, for some clients, formal assessment strategies have an influence above and beyond a counselor's verbal comments. For these clients, assessment results either presented on a computer screen or printed on paper have greater credibility than what they know themselves or what the counselors may be telling them. For example, I have worked with career counseling clients who could articulate career preferences but who did not see those interests as being legitimate until the results of an interest inventory confirmed their preferences. In addition, a counselor can quickly tarnish her reputation by not being adequately prepared in assessment. As an example, a school counselor can quickly lose credibility with a student's parents by not being able to interpret the results from an achievement test.

LO2 Assessment Is Integral to Counseling

Some counseling students may view assessment as using only formal instruments and see it as an activity distinct from the counseling process. Consider for a moment the counseling process and the essential steps included in it. Although the counseling process is quite complex, it encompasses the following four broad steps:

1. Assessing the client problem(s)
2. Conceptualizing and defining the client problem(s)
3. Selecting and implementing effective treatment(s)
4. Evaluating the counseling

The following is a discussion of how assessment is an integral part of the counseling process. Sometimes clinicians think of assessment only as formal tests that are expensive and time consuming. I have a broader view of assessment that is focused on gathering information holistically and using different assessment tools throughout the counseling process.

Assessing the Client Problem(s)

In the initial stage of counseling, a counselor needs to assess the client problem because there is no "one-size-fits-all" approach to the therapeutic process. It is important that counselors skillfully assess the client's problem, for if the assessment process is incomplete or inaccurate, the entire counseling process can be negatively affected. Furthermore, if counselors have limited assessment skills, they may miss or underestimate important client issues. If problems are delineated in an efficient manner, treatment can be initiated sooner.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Also, in mental health settings, unlike other counseling settings such as schools, counselors need to diagnose the client's problem in order for third-party payers (e.g., government, commercial insurers, managed care organizations) to pay for the mental health services (Gersten, 2013). More information on diagnosis will be provided in Chapter 15.

Epperson, Bushway, and Warman (1983) found that clients are more likely to continue in counseling if they and the counselor agree on the nature of the problem. Hood and Johnson (2007) recommended combining different types of assessment data (e.g., formal and informal) to maximize the strengths and minimize the limitations of different strategies. Meyer et al. (2001) found substantial empirical support for the practice of using multimethod assessment batteries and concluded that clinicians who use interviewing exclusively often have incomplete client information. Assessment skills, however, are not needed solely in this first stage of counseling; they are important throughout the entire therapeutic process.

Conceptualizing and Defining the Client Problem(s)

A counselor may be extraordinarily skilled at initial assessment, but if she refrains from gathering additional information, then the process will be hampered. During conceptualizing, the second stage of the process, counselors need to continually assess a client to ensure that they maintain an adequate understanding of the client's needs and problems. Distinguishing between simple and complex problems is critical in the selection of treatment and the effectiveness of the counseling. Mohr (1995) found that one of the most robust predictors of negative outcomes in counseling and psychotherapy is the clinician underestimating the severity of the client's problems.

Dawes (1994) presented some compelling evidence that counselors are not always objective in their perceptions and analyses of client issues. There is research supporting the notion that counselors do have a tendency toward *confirmatory bias* (Spengler, Strohmer, Dixon, & Shivy, 1995), which entails the tendency to seek evidence that confirms an individual's preferred hypothesis. For example, if a counselor believes that substance abuse is rampant in our society, that clinician may have a tendency to perceive many clients' issues as being related to their abuse of substances. Ægisdóttir et al. (2006) conducted a meta-analysis of studies that compared the predictions of mental health professionals with statistical methods and found that statistical methods were slightly more effective in making clinical decisions than mental health clinicians.

Fredman and Sherman (1987) proposed that clients often benefit from formal assessments because they can provide counselors with a different avenue for reaching the client. They contended that counselors should occasionally get away from relying completely on auditory experience and add visual, kinesthetic, and tactile dimensions to their sessions. They suggested that testing can provide a visual experience that often inspires more confidence in the information than do spoken words. In addition, this visual experience often motivates clients to take action about a conflict or problem. Thus, I assert that assessment skills are necessary for adequately conceptualizing the client's concerns, identifying contextual factors that may be contributing to the problems, and ascertaining factors that may be helpful in the treatment phase of counseling.

Selecting and Implementing Effective Treatment(s)

The third step in counseling involves selecting and implementing treatment that is based on the previous assessments; however, assessment does not stop once clinical treatment begins. Rather, the assessment process continues throughout the treatment phase of counseling. This is the stage of counseling where a counselor might use either informal or formal assessment to answer a therapeutic question or monitor progress.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Clients often come to counseling for assistance in answering questions such as: Should I drop out of school? Should I get a divorce? Should I change careers? Increasing the amount of information a client can use in decision making usually enhances his or her decision-making process. Selecting instruments that generate information about occupational choices, career interests, and family dynamics can often assist clients in making major decisions. Using assessment information to aid clients in decision making, however, does not always necessitate administering additional tests. Counselors can often use assessment instruments that clients have taken during previous educational experiences or in other situations. As an example, one client, who had taken the American College Testing (ACT) exam two years earlier, made his decision to pursue a new job after reviewing his scores from that test. Campbell (2000) argued that tests in counseling should provide new information and that counselors can use assessment results to encourage client learning and insight.

Sometimes clinicians believe that assessment is used only to identify problems or pathologies. This seems somewhat inconsistent with counseling's developmental and preventative foundation. Rather than using tests to diagnose pathologies or identify limitations, counselors can use an instrument's assessment to reveal strengths. Lopez, Snyder, and Rasmussen (2003) argued that clinicians must strike a vital balance between assessing strengths and limitations. Drawing from the substantial research supporting positive psychology, these authors contended that psychological assessment is currently slanted toward identifying limitations, even though there is compelling empirical evidence that constructs such as hope, resiliency, learned optimism, and courage are intimately tied with well-being and performance. Hence, if practitioners want to build on positive processes, outcomes, and environments, they should also assess the degree to which these positive factors are present. Wright and Lopez (2002) suggested a four-pronged approach to positive psychological assessment. In this approach, counselors should identify (1) undermining characteristics of the client, (2) client strengths and assets, (3) missing and destructive aspects of the client's environment, and (4) resources and opportunities within the environment. Counselors can also use positive psychological assessment to chart changes during the counseling process and to measure the development of client strengths.

In addition, during the third step in counseling, counselors need to continually monitor whether they have comprehensively understood the client and the client's situation. Furthermore, practitioners need to scrutinize whether progress is being made and adjust the counseling process if the client is not making progress toward the therapeutic goals. Research indicates that better client outcome is associated with therapists systematically and consistently monitoring client progress throughout the entire counseling process (Lambert, 2010).

Evaluating the Counseling

Finally, once the treatment phase is completed, counselors need to assess or evaluate whether it was effective. Once again, just as counselors need effective communication skills to help clients, they also need effective assessment skills related to evaluation. Outcome assessment can provide clinicians information about their own effectiveness and, if gathered by all clients, can provide information to administrators, appropriate boards (e.g., a school board), and other relevant organizations (e.g., a governmental organization that controls grant funding).

In the counseling profession, clinicians are frequently more interested in helping individuals than in showing that the services they provide are effective. This approach, however, puts many clinicians at risk, because there are increasing demands from legislators, administrators, foundations, managed-care providers, clients, and parents for tangible

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

documentation showing that counseling is both helpful and cost-effective. In these times of budgetary constraints, when school counseling positions could be cut, school counselors need to have accountability information readily available to document their usefulness (American School Counselor Association [ASCA], 2012). Watson and Flamez (2015) documented the necessity for community agencies and health care organizations to have well-designed accountability systems. Not only do counselors need knowledge of assessment to meet the accountability demands but outcome information can also be used in the counseling process. Some clients finding it very empowering to view assessment results that indicate they have made positive changes (e.g., feel less depressed).

Assessment Can Be Therapeutic

The view that assessment can be used in a therapeutic manner is sometimes contrary to how professionals think of testing and assessment. Counselors can, however, use assessment results to stimulate clients' self-awareness and exploration. Finn (2007) suggested a model in which the assessment itself is considered to be a therapeutic intervention. This *therapeutic assessment model* is compared with the traditional *information-gathering model*, in which assessment is viewed as a way to collect information that will guide subsequent treatment. The intent of therapeutic assessment is to promote positive changes in clients through the use of assessment instruments.

As will be discussed in Chapter 5, ethical guidelines state that clients have the right to assessment results; thus, therapeutic assessment recommends that the results be provided in a collaborative and therapeutic manner. In a meta-analysis of studies of therapeutic assessment, Poston and Hanson (2010) found that those who received a therapeutic style of assessment results had much better outcomes than those who received no assessments or a traditional model of providing assessment results.

Michel (2002) provided some examples of using assessments therapeutically with clients who have eating disorders. She argued that for these clients, assessments can provide self-verifying information, which can be very therapeutic. In addition, she indicated that "joining" or aligning with the clients' families can be facilitated by providing assessment information that acknowledges the difficulties of their situation. She recommended the following assessment strategies:

1. Continue to nurture and maintain the therapeutic alliance throughout the evaluation.
2. Present all findings in a helpful context, emphasizing that the results reflect the patient's communication.
3. Recognize that defensive protocols are useful, and attempt to discern the reasons for such responding.
4. Encourage and allow the patient to respond affectively, and give/modify examples of findings.
5. Tie the findings to the adaptive functions of the eating disorder.
6. Use examples from the assessment process and test findings to provide education about eating disorders to help normalize the patient's feelings.
7. Use psychological test results in the same manner with family members to enlist family involvement. (p. 476)

In counseling, assessment results can also be used in a therapeutic manner by helping clients make effective decisions. Counselors can use information gained from assessment to help clients weigh the pros and cons of a decision and examine the probabilities of expected outcomes. This manner of using assessment information, which is common in career counseling, can also be used with assessment results related to other issues.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L03 What Do Counselors Need to Know About Assessment?

Because assessment is an integral part of counseling, it is crucial that practitioners become competent in this area. CACREP (2009) accredits many counseling programs and has published standards, which include specific standards related to assessment. The following are the standards for assessment and a brief explanation of where the content related to these standards is addressed in this book:

> *ASSESSMENT—studies that provide an understanding of individual and group approaches to assessment and evaluation in a multicultural society, including all of the following:*

A. historical perspectives concerning the nature and meaning of assessment;

A historical overview of assessment is provided later in this chapter. Although some readers may find a history of assessment a little tedious, it is important to know about the evolution of assessment and learn from some of the past mistakes that have occurred in this field.

B. basic concepts of standardized and nonstandardized testing and other assessment techniques, including norm-referenced and criterion-referenced assessment, environmental assessment, performance assessment, individual and group test and inventory methods, psychological testing, and behavioral observations;

Basic concepts regarding assessment techniques start in Chapter 2, which addresses the differences between norm-referenced and criterion-referenced assessments for both standardized and nonstandardized testing. Although this book focuses more on individual and group assessments of clients, it also includes some information on environmental assessments. In particular in Chapter 14, I talk about assessing the family environment, which is an important environmental influence for many individuals. Throughout the book, however, I stress that it is important to consider environmental influences on the individual when conducting any type of assessment. The definition of performance assessment is included in the next section of this book, and trends in this area are discussed particularly as related to achievement testing and assessment of intelligence. Furthermore, the middle section of this book is devoted to individual and group inventories, which includes psychological testing and behavioral observations.

C. statistical concepts, including scales of measurement, measures of central tendency, indices of variability, shapes and types of distributions, and correlations;

An overview of statistical concepts relevant to assessment begins in Chapter 2 and is continued in Chapters 3 and 4. The thrust of this discussion is not to make the reader a statistical wizard, but to focus on statistical concepts that are necessary and useful in assessing clients and in interpreting results. For the math phobic reader, please focus on the statistical concepts as compared to the calculations. Many of the calculations, however, are quite simple and designed to help the reader understand the statistical concepts. Scales of measurement, measures of central tendency, indices of variability, and shapes and types of distributions are all covered in Chapter 2. Correlation is introduced in Chapter 3 and discussed again in Chapter 4.

D. reliability (i.e., theory of measurement error, models of reliability, and the use of reliability information);

Reliability is an important concept in assessment and it is addressed in detail in Chapter 3. Included in Chapter 3 is a discussion of classical test theory or theory of measurement error. Also covered are models of reliability and different methods for calculating estimates of reliability. In addition, I discuss how to use reliability information in both evaluating instruments and interpreting results to clients.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

E. validity (i.e., evidence of validity, types of validity, and the relationship between reliability and validity);

Validation evidence informs the clinician on how the instrument can be used and what the results mean; therefore, it is a critical aspect of assessment. Chapter 4 of this book includes information on the relationship between reliability and validity and the different types of validation evidence typically provided. When using formal assessments, counselors should focus on the reliability and validity evidence provided in the instrument's manual. When using informal assessments, counselors should also consider reliability and validity while using these instruments as reliability and validity evidence is not just applicable to formal instruments.

F. social and cultural factors related to the assessment and evaluation of individuals, groups, and specific populations;

As the demographics of the United States are rapidly changing, it is important for counselors to consider relevant social and cultural factors. This is particularly important in the area of assessment where instruments may have been developed for appropriate use with only select clients. Therefore, counselors must be aware of social and cultural factors in assessment as interpretation may need to be more tentative with a client with a specific racial or ethnic background. Multicultural issues in assessment are frequently addressed in this book; however, these issues are specifically discussed in Chapter 6.

G. ethical strategies for selecting, administering, and interpreting assessment and evaluation instruments and techniques in counseling.

Ethics are standards that guide the behaviors of professionals. In the ACA *Code of Ethics* (2014), an entire section is devoted to the area of evaluation, assessment, and interpretation (i.e., Section E). This topic is so important that this entire section of the *Code of Ethics* has been reproduced in Appendix A of this book.

LO4 Types of Assessment Tools

In order to begin to understand the use of assessment in counseling, a practitioner needs to understand some of the basic types of assessment instruments. Assessment and testing are topics that often arise when counselors consult with other professionals (e.g., psychologists, social workers, and teachers). The field of mental health is moving toward a multidisciplinary team approach to treatment; hence, if counselors want to continue to be a part of this multidisciplinary team, they need to understand the nomenclature of assessment. Although there are many different ways to classify assessment instruments, the following information provides an overview of some commonly used categories and terms.

Standardized vs. Nonstandardized

For an assessment device to be considered a standardized instrument, there must be fixed instructions for administering and scoring the instrument. In addition, the content needs to remain constant and to have been developed in accordance with professional standards. If the instrument is comparing an individual's performance with that of other individuals, the instrument must be administered to an appropriate and representative sample. A nonstandardized instrument has not met these guidelines and may not provide the systematic measure of behavior that standardized instruments provide.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Individual vs. Group

This distinction concerns the administration of the instrument. Some instruments can be given to groups, which is often a more convenient and less time-consuming method of administering an instrument as compared to administering it to one person at a time. With group administration, however, it is difficult to observe all the examinees and to note all of their behaviors while they take the instrument. A substantial amount of information can often be gained by administering an instrument individually and by observing a client's nonverbal behaviors. Some well-known psychological instruments are administered only individually in order to gather relevant clinical information.

Objective vs. Subjective

This categorization reflects the methods used to **score** the assessment tool. Many instruments are scored objectively; that is, there are predetermined methods for scoring the assessment, and the individual doing the scoring is not required to make any judgments. Subjective instruments, on the other hand, require the individual to make professional judgments in scoring the assessment. For example, many multiple-choice tests are objective instruments, with the scoring completed by noting whether the test-taker's response is correct or incorrect. Essay tests, however, are usually subjective instruments because the person grading these exams must make some judgments about the quality of the answers. Objectively scored instruments attempt to control for bias and inconsistencies in scoring. However, counselors are often interested in exploring clients' issues, which are not easily assessed by using only objective methods.

Verbal vs. Nonverbal

In recent years, counselors have become increasingly aware of the influences of language and culture on assessment. Instruments that require examinees to use verbal skills can be problematic for individuals whose primary language is not English. Imagine that you are being tested and you read the following instructions: *Da prodes test ti moras zavoriti ovu knjigu i zafpati*. Many people would fail this test because they would not be able to read the Bosnian instructions that say, "In order to pass this test, you must close this book and take a nap." (I know that some readers would be more than willing to oblige with these instructions now that they understand the translation.) Even when a test does not involve verbal skills, if the instructions are given orally or must be read, it is still considered a verbal instrument. Some people prefer the term *nonlanguage* instead of *nonverbal* to denote instruments that require no language on the part of either the examiner or the examinee. Another term related to this topic is **performance tests**, which require the manipulation of objects with minimal verbal influences (e.g., putting a puzzle together, arranging blocks in a certain design). Sometimes there is not a clear distinction between a verbal and a nonverbal test. It is difficult to design instruments that have no language or verbal components. Hence, with some clients, the counselor may need to determine the degree to which language and verbal skills influence the results. In multicultural assessment, the practitioner needs to consider the degree to which both culture and language influence the results.

Speed vs. Power

This classification concerns the difficulty level of the items in an assessment. In power tests, the items in the examination vary in difficulty, with more credit given for more difficult items. A speed test simply examines the number of items completed in a specified time period. The determination of whether an assessment is a speed or a power test depends on the purpose of the assessment. If the aim of an assessment is to determine how quickly people can do a specific task, then a speed test is appropriate. If determining the mathematical abilities of an individual is the goal, then a power test is needed.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Cognitive vs. Affective

Cognitive instruments are those that assess cognition: perceiving, processing, concrete and abstract thinking, and remembering. Three types of cognitive tests predominate: intelligence or general ability tests, achievement tests, and aptitude tests. **Intelligence tests** are sometime called **general ability tests** because the term *general ability* does not have the same connotation as *intelligence*. Intelligence/general ability instruments typically measure a person's ability to think abstractly, solve problems, understand complex ideas, and learn new material—abilities involved in a wide spectrum of activities. Although intelligence is a complex phenomenon, it is essentially related to how "smart" the individual is (Sternberg, 1985). **Achievement tests**, which are measures of acquired knowledge or proficiency, measure the extent to which an individual has "achieved" in acquiring certain information or mastering certain skills. As an example, after individuals have gone through instruction or training, they are often assessed to determine how well they acquired the knowledge or skill. Classroom tests of a single academic subject are the most common form of achievement tests. Whereas achievement tests measure whether an individual has acquired some knowledge or skill, **aptitude tests** predict an individual's performance in the future. Achievement tests and aptitude tests can be similar in content, but their purposes are different. Aptitude tests do not measure past experiences; rather, they assess an individual's ability to perform in the future—to learn new material or skills.

Affective instruments assess interest, attitudes, values, motives, temperaments, and the noncognitive aspects of personality. Both informal and formal techniques have a dominant role in affective assessment. In the area of formal instruments, practitioners most frequently use one of two types of personality tests: structured instruments and projective techniques. **Structured personality instruments** include the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), in which individuals respond to a set of established questions and select answers from the provided alternatives. With **projective techniques**, individuals respond to relatively ambiguous stimuli, such as inkblots, unfinished sentences, or pictures. The nonstructured nature of the projective techniques provides the examinee with more latitude in responding. It is theorized that these nonstructured responses are projections of the individual's latent traits. Although projective techniques are often more difficult for the examinee to fake, the examiner needs to be extensively trained to use these instruments appropriately. This book does not provide the required information and supervision that are necessary for using projective techniques.

L05 History

To understand current assessment techniques and instruments, counselors need information about the history of assessment. The goal of this brief excursion into the development of assessment is to provide a context for understanding the current state of the field. Through knowledge of relevant issues and with an understanding of why and how some instruments were developed, counselors can begin the process of becoming informed, competent users of assessments.

Early Testing

Assessment is not a new phenomenon; testing has been around for many centuries (Anastasi, 1993; Bowman, 1989). There is some evidence that the Greeks might have used testing approximately 2,500 years ago. The Chinese used a civil services examination 2,000 years ago, and even then, there were discussions of the effects of social class, cheating, and examiner bias (Bowman, 1989). In general, however, the English biologist Francis Galton is credited with launching the testing movement. Galton did not set out to initiate testing, but in his study of human heredity during the late 1800s and early 1900s, he wanted a way to measure human characteristics of biologically related and unrelated individuals. Galton believed that

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

he could use sensory discrimination tests to measure individual intelligence. He based this opinion on the premise that all information is conveyed through the senses, and thus, the more perceptive a person is, the greater the amount of information that is accessible for intelligent judgments and actions. Galton's book *Hereditary Genius* was published in 1869. Galton also made a significant contribution in the area of statistics. The commonly used statistical technique of correlation came from his work in heredity and "co-relations."

Another psychologist who influenced early assessment was Wilhelm Wundt. Wundt, the man credited with founding the science of psychology, was also interested in measuring psychological constructs. Wundt's purpose was somewhat different from Galton's in that Wundt and his colleagues were more interested in identifying factors of intelligence that were common or innate to all human beings. Another prominent figure in the early testing movement was the American psychologist James McKeen Cattell. Cattell was Wundt's student. Drawing from the work of both Galton and Wundt, Cattell expanded testing to include memory and other simple mental processes. Cattell was the first to use the term *mental test*, although his mental tests were quite simple in that they were not strongly related to estimates of school achievement or to other criteria now considered indicative of intelligence (Anastasi, 1993).

1900–1920

Just prior to the turn of the century, Binet and Henri published an article criticizing most available intelligence tests as being too sensory oriented. Binet was given the opportunity to develop his ideas when the French Minister of Public Instruction appointed him to a commission studying "educationally retarded" children. The result was the first version of the Binet–Simon scale, published in 1905. The instrument, which was individually administered, was a simple instrument with only 30 items and a norming group of 50. The Binet–Simon scale was different from previous measures of intelligence in that it focused on assessing judgment, comprehension, and reasoning. The instrument was revised in 1908 to incorporate a ratio of mental age level to chronological age level. This was labeled the *intelligence quotient (IQ)*. This method of calculating IQ was used for many years, although, as will be discussed in later chapters, there are some problems associated with it. Binet's instrument was translated and adapted in several countries, including an adaptation by L. M. Terman at Stanford University. The revised test became known as the Stanford–Binet scale; however, the official title of Terman's 1916 publication is *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet–Simon Intelligence Scale*. Although other translations of the Binet–Simon scale were available at that time, Terman's normative sample and his methodical approach were credited with the success of the Stanford–Binet (Becker, 2003).

While individually administered intelligence tests were being developed, there was also some interest in group testing, particularly by military leaders in the United States during World War I. In 1917, the army contacted Robert Yerkes, then president of the American Psychological Association, to have psychologists assist in developing a group intelligence test. The army hired psychologists and a group-administered intelligence assessment was developed for use in selection and classification of personnel. The army used a multiple-choice format, which had only recently been introduced by Arthur Otis. These first group-administered intelligence tests were known as Army Alpha, which was used for routine testing, and Army Beta, which was a nonlanguage instrument designed for use with illiterate or non-English-speaking recruits (Figure 1.1). Although these instruments were never really used in World War I, shortly after the end of the war, the army released these two instruments for public use.

Another influential person of this time was Frank Parsons, who is often cited as being the father of guidance. Although Parsons did not develop any assessment, he did devise one of the first approaches to career counseling. In his three-step career counseling

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

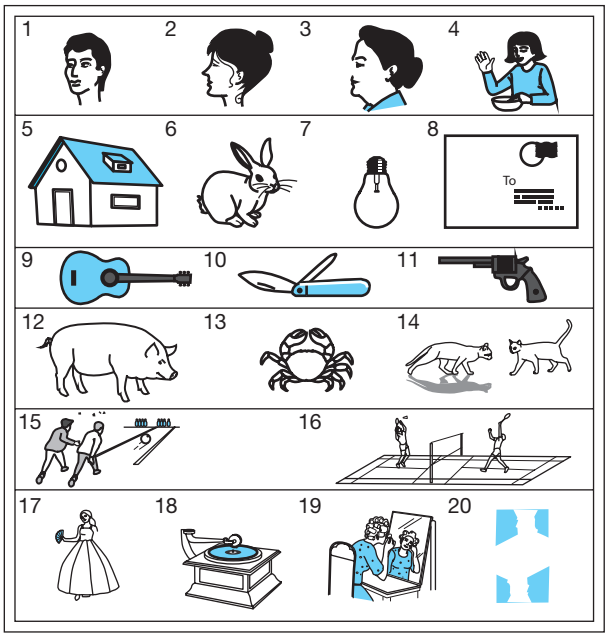


FIGURE 1.1
Answers: 1) Mouth; 2) Eye; 3) Nose; 4) Hand; 5) Chimney; 6) Ear; 7) Filament; 8) Return address; 9) Strings; 10) Corkscrew 11) Trigger; 12) Tail; 13) Claw; 14) Shadow; 15) Ball; 16) Net; 17) Arm; 18) Speaker; 19) Arm in mirror; 20) Diamond
Retrieved from http://official-asvab.com/armysamples_res.htm

model, a counselor would (1) understand the person, (2) understand the world of work, and (3) match the person to the appropriate occupation in the world of work. Some career counselors have accomplished this first step of gaining an understanding of the person by incorporating some of the career assessments that have been developed since Parsons' time.

1920s and 1930s

Interestingly, the first intelligence tests were developed without a theoretical foundation, but in the 1920s and 1930s there was considerable theoretical attention to defining and explaining the construct of intelligence (Erford, Moore-Thomas, & Linde, 2007). Spearman (1927) proposed that intelligence consisted of two types of factors—one that pertained to general tasks and another that pertained to specific tasks. Thurstone (1938), on the other hand, proposed that there is no one general factor of intelligence; rather, there are seven primary mental abilities. This debate about whether intelligence is one general factor (*g*) or multiple factors continues, and even today, theoreticians have differing views of intelligence. During this time, there also was increased activity related to developing measures of intelligence. Terman and Merrill revised the 1916 version and the test was published using its current name, the Stanford–Binet. In 1939, another significant test, the Wechsler–Bellevue Intelligence Scale, was published. The primary author was David Wechsler, who continued to make significant contributions to intelligence testing until his death in 1981.

Interest in assessment was not restricted to intelligence testing alone. At the end of World War I, there was also an interest in identifying men who were not emotionally capable of serving in the armed forces. The army once again developed the prototype, but this time the group test, called the Woodworth's Personal Data Sheet, focused on personality assessment. Consistent with the Army Alpha and Army Beta, this self-report inventory was released for civilian use, where it spurred the development of other self-report personality inventories. In 1921, Rorschach described the technique of using inkblots as a tool for diagnostic investigation of the personality. Although his methods did not have a significant impact immediately, his writings had a major influence on clinical assessment later.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Another projective test of personality (i.e., Thematic Apperception Test) was published in 1935. In this assessment, Murray and Morgan presented individuals with pictures and the examinees told stories about the pictures. The examiner would then evaluate the text in order to identify themes. A more detailed discussion of both of these projective assessments is included in Chapter 12.

As private industries began to see that tests could be used for selecting and classifying industrial personnel, special aptitude tests were developed, primarily for use in clerical and mechanical areas. There was also the development of vocational counseling instruments, such as the publication in 1927 of the Strong Vocational Interest Blank by E. K. Strong. Another individual active in the development of interest inventories for vocational counseling was G. Frederic Kuder, who first published the Kuder Preference Record—Vocational in 1932.

In 1923, the first standardized achievement battery, the Stanford Achievement Test, was published. This instrument was designed to provide measures of performance in different school subjects as opposed to testing in only a single subject. These achievement batteries also allowed educators to compare students' performances to other students from various parts of the country. In the mid-1800s, there was a move from oral achievement examinations to written essay testing. By the 1930s, there was considerable evidence concerning the difficulties of written essay examinations, particularly due to the lack of agreement among teachers in grading essay items. The desire for more objectivity in testing promoted the use of more objective items and the development of state, regional, and national achievement testing programs.

The rapid advancement in many areas of testing during the 1920s and 1930s led to a need for a resource to identify and evaluate testing instruments. The first edition of the *Mental Measurements Yearbook* was published in 1939 to fill that need. Oscar Buros established these yearbooks to provide information about instruments as well as to critique the properties of the instruments. The Mental Measurements Yearbooks continue to be a good resource for counselors currently.

1940s and 1950s

Even though there were efforts to produce sound personality assessment tools, there was some dissatisfaction with many of these instruments because they were somewhat transparent and could easily be faked by an examinee. Therefore, projective techniques, such as the Rorschach, began to become more popular. With projective techniques, clients respond to a relatively unstructured task, and their responses are then evaluated. The increased use of projective techniques did not, however, hamper the development of self-report instruments. In the early 1940s, Hathaway and McKinley developed a prominent personality instrument, the Minnesota Multiphasic Personality Inventory (MMPI), which incorporated validity scales to assess the degree to which individuals portrayed themselves in an overly positive or negative way. The MMPI also contained items that were empirically selected and keyed to criterion rather than items that appeared to measure different aspects of personality.

Standardized achievement tests were also becoming well established in the public schools. Although single aptitude tests were already in existence, most of the multiple aptitude batteries appeared after 1940. Multiple aptitude batteries could indicate where an individual's strengths and limitations were (e.g., Does an individual have higher verbal aptitude or higher numerical aptitude?). The development of multiple aptitude batteries came late compared with other assessment areas, which was directly related to the refinement of the statistical technique of factor analysis.

With the increased use of assessment instruments, problems associated with these instruments began to emerge. As criticisms of assessment rose, it became clear that there was a need for standards with respect to the development and use of instruments. The American Psychological Association published the first edition of defined standards.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In later years, three organizations—the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME)—collaborated on revised editions, titled *Standards for Educational and Psychological Testing.* These standards continue to be revised, with the most recent edition published in 2014, and serve as a significant resource in the evaluation and appropriate use of appraisal instruments.

As assessment became more established, it also became more sophisticated. Individuals began to see that centralized publication of tests would be convenient for consumers as well as, quite possibly, profitable for the publishers. For example, the testing functions of the College Entrance Examination Board (CEEB), the Carnegie Corporation, and the American Council on Education merged to form the Educational Testing Service. With the centralization of some publishing, electronic scoring became more cost-effective. Electronic scoring reduced scoring errors and allowed for more complicated scoring procedures.

1960s and 1970s

This period is marked by an examination and evaluation of testing and assessment. The proliferation of large-scale testing in schools, along with the increased use of testing in employment and the military, led to widespread public concern. Numerous magazine articles and books questioned the use of psychological instruments and uncovered misuses of these instruments. Assessment instruments were particularly scrutinized for racial or ethnic bias, fairness, and accuracy. This scrutiny revealed many limitations of existing instruments, particularly those concerning the use of some instruments with minority clients.

An interesting paradox occurred during the 1970s. Although there was substantial concern about the use of tests, there was also a grassroots movement related to minimum competency testing that encouraged more testing. This movement grew out of concern that high school students were graduating without sufficient skills. Because the public wanted to ensure that children reached a minimal level of competency before they graduated, many states enacted legislation requiring students to pass a minimum competency examination before they could be awarded a high school diploma (Lerner, 1981). In addition, legislation at the national level had a significant impact on testing. In 1974, the Family Educational Rights and Privacy Act was passed, mandating that parents and children older than 18 had the right to review their own school records. This legislation also specified topics that could not be assessed without parental permission.

As the country was beginning to take advantage of advances in technology, the field of assessment was also changing. Computers began to be used more, particularly for scoring. Near the end of the 1970s, the assessment field began to explore the interactive nature of computers and methods for using computers not only to score but also to administer and interpret assessment results

1980s and 1990s

The use of computers in appraisal really blossomed in the 1980s. With the increased availability of personal computers, clients could take an assessment instrument on a computer and receive the results immediately. Computers could be programmed to adapt the order of items depending on a client's previous answer. For example, if a client got one item correct, then the next item could be more difficult, but if a client got the item wrong, the next item could be easier. There also was an increase in computer-generated reports, where instead of a psychologist writing a report, the report was written by a computer.

Earlier criticisms of assessment techniques led to many instruments being revised. For example, the developers of the MMPI-2 attempted to eliminate many of the shortcomings of the MMPI. The most widely used instrument to assess children's intelligence, the Wechsler Intelligence Scale for Children—Revised (WISC-R), was revised to become

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the WISC-III. The adult version of the Wechsler scale (WAIS) was revised twice during this period. New instruments designed to be sensitive to cultural diversity were also developed (e.g., Kaufman Assessment Battery for Children). Issues related to cultural bias and sensitivity to multicultural influences that arose in the 1960s and 1970s continued to be researched and discussed. Many professional organizations began to realize that standards for multicultural counseling and assessment needed to be developed.

A major testing movement within the 1990s was authentic assessment, the purpose of which is to evaluate using a method consistent with the instructional area and to gather multiple indicators of performance. Authentic assessment has had a major influence on teachers' assessments of students' academic progress (Fisher & King, 1995). Teachers often use portfolio assessment, where multiple assignments are gathered together and evaluated. Rather than using only one multiple-choice test, teachers evaluate multiple essays, projects, and tests that are designed to represent the material being taught.

2000 to the Present

It is difficult to predict the future issues and trends that will evolve in assessment. Technology and the Internet are continuing to change the manner in which assessments are developed, administered, scored, and interpreted. Counselors can now have clients sit down at a computer and take many assessments through the publisher's website. There are also some drawbacks to technology's influence on assessment. Thousands of assessments are on the Internet; however, it is difficult for clients to determine what assessments have been researched and the degree to which validation evidence has been accumulated. Individuals might make important life decisions (e.g., whether to pay for an expensive training program) based on an instrument that has very low reliability.

Multicultural issues will continue to be a focus of research in the assessment area. Identifying methods of assessing individuals with different cultural backgrounds is complex, and it is hoped that the field will make strides in this area. A number of instruments have already been revised since 2000 with the goal of diminishing instrument bias. In addition, as we progress toward a global society, there is interest in assessment instruments that can be used in multiple countries.

Certainly the passage of the No Child Left Behind Act of 2001 has had a dramatic influence on achievement testing in the schools. The specifics of No Child Left Behind related to assessment will be discussed in Chapter 10. In many ways, the Act was passed to make schools more accountable. The theme of accountability is also evident for counselors interested in being mental health counselors. Indications are that counselors will need to be more accountable and provide effectiveness data to various constituencies. In particular, the ASCA developed the ASCA National Model (ASCA, 2012), which encourages school counselors to collect and use data that link school counseling programs to student achievement and other important indicators of effectiveness.

Arguably, this is another time of change within the field of assessment. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), the primary resource for assessment, was recently revised. Its purpose "is to provide criteria for the development and evaluation of tests and testing practices and to provide guidelines for assessing the validity of interpretation of test scores for the intended test uses" (p. 1). Its revision provides more clarity about recent advancements in the assessment and focuses on test fairness. Another major revision is of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), with the publication of the *DSM-5* (American Psychiatric Association, 2013). The American Psychiatric Association published the first edition of the *DSM* in 1952 to provide a common nomenclature regarding mental disturbances and disorders. Through various revisions, the *DSM* has served as the guidebook for the diagnosis of mental disorders for

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

many mental health clinicians. There are numerous changes in the *DSM-5* as compared to other editions, and an overview of the *DSM-5* is provided in Chapter 15. Other future trends in assessment may turn out to be significant, but what is clear is this is an exciting time for counselors with regard to assessment.

Summary

Often the terms *assessment* and *appraisal* have different meanings in the helping profession, depending on the situation and context. This book defines both terms as procedures, including formal and informal techniques, designed to gather information about clients. Assessment can involve formal tests, psychological inventories, checklists, scales, interviews, and observations. The degree to which counselors accurately assesses clients will influence the effectiveness of their counseling. Assessment is an integral part of the counseling process, and as such, counselors need to develop proficient assessment skills.

The history of assessment in counseling indicates that there has been an ebb and flow in assessment strategies. In the United States today, there is considerable interest in various areas of testing, and some environments, such as schools, have seen a significant increase in assessments. Given current social and political influences, it is even more important for counselors to understand issues in assessment. The development of assessment skills involves knowledge of basic measurement techniques and methods for evaluating instruments. These skills also involve knowledge of the constructs (e.g., intelligence, personality) being measured and how to use the results in counseling. Assessment results can sometimes have a dramatic effect on a client; therefore, counselors should strive to ensure that assessment results are used in a professional and ethical manner. In the past, psychologists have challenged whether counselors have the appropriate training and background to use assessment. If counselors want to continue to have a place in providing mental health services, then they need to be proficient in assessment. The following chapters are designed to provide counselors with an introduction to key issues in assessment with an emphasis on understanding basic concepts related to measurement.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 2

Basic Assessment Principles

Learning Objectives

- **LO1** Distinguish among the measurement scales
- **LO2** Differentiate between norm-referenced and criterion-referenced instruments
- **LO3** Construct a frequency distribution
- **LO4** Describe three measures of central tendency
- **LO5** Calculate variance and standard deviation
- **LO6** Understand the characteristics of a normal distribution
- **LO7** Interpret a percentile score
- **LO8** Distinguish among *z* scores, *T* scores, and stanines
- **LO9** Understand the different methods for gathering a normative sample

Let's assume that you have just taken a test called the Counseling Aptitude Scale (CAS), and you received a score of 60 on this fictitious instrument. It is difficult to draw any conclusions based solely on this score of 60. For example, we don't know if the scores range from 0 to 60 or if the range is from 60 to 600. Even if the scores range from 0 to 60, a score of 60 would mean something different if everyone else taking the instrument had a score of 58 than it would if everyone else scored a 30. Any score on any instrument is difficult to interpret without additional information. The basic principles discussed in this chapter will assist counselors in understanding assessment scores and results and will help them begin to know how to interpret those scores to clients. In fact, I would argue that a test user without knowledge of basic measurement principles is analogous to a tourist lost in a strange city without any navigation assistance.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

LO1 Measurement Scales

When counselors use assessments, they are attempting to *measure* some aspect of the client. Measurement typically involves the application of specific procedures for assigning numbers to objects. When an entity is measured, there are rules for how the measurement is performed. As an example, there is a rule related to the length of an inch. An inch is always the same distance regardless of the place or time that it is measured. In measuring any matter—a person's weight, the temperature outside, the number of lawyers in New York City, or the emotional stability of graduate students—the rules for measuring depend on the precision of the measurement scale being used.

There are four basic types of measurement scales: nominal, ordinal, interval, and ratio. Any measurement of a client (whether it is height or level of anxiety) involves one of these four types of measurement scales. Numbers at different levels or scales of measurement convey different types of information (Cohen, Swerdlik, & Sturman, 2013); therefore, counselors need to determine the type of measurement scale being used in an assessment in order to evaluate the instrument properly. Statistical tests are often used to analyze the psychometric qualities of an instrument. The type of measurement scale will influence the selection of appropriate statistical techniques. Therefore, one of the early steps in evaluating an instrument is to identify the measurement scale used and then determine if the statistical analyses are appropriate for that type of measurement scale.

The **nominal scale**, which is the most elementary of the measurement scales, involves classifying by name based on characteristics of the person or object being measured. In other words, the intent of the nominal scale is to name an object. If this scale is used, there is no indication of amount or magnitude. An example of a nominal scale is 1 = Democrats, 2 = Republicans, 3 = Libertarians, 4 = No party affiliation. When numbers are assigned to groups, it is not possible to perform many of the common statistical analyses. For example, the average for the above numbers for party affiliation is 2.5, which does not convey any real meaning. With a nominal scale, we can get only a count or a percentage of individuals who may fall into a specific category.

An **ordinal scale** provides a measure of magnitude and, thus, it often provides more information than does a nominal scale. An instrument with an ordinal scale makes it possible to determine which scores are smaller or larger than other scores. Ranking a group of children from best to worst behaved in the classroom is an example of an ordinal scale. The ordinal scale enables one to rank or order individuals or objects, but that is the extent of the precision. On the fictitious Counseling Aptitude Scale mentioned earlier, you received a score of 60, whereas Jason's score was 50 and Rebecca's was 90. Unless we know that there are equal intervals between points throughout the scale, all we can say is that Rebecca received the highest score, you were in the middle, and Jason had the lowest score of the group. With any kind of measurement, unless there are equal units (e.g., inches have equal units), the degree to which there is a difference between the scores cannot be precisely determined.

In an **interval scale**, the units are in equal intervals; thus, a difference of 5 points between 45 and 50 represents the same amount of change as the difference of 5 points between 85 and 90. It is often difficult to design instruments that measure psychological and educational aspects in equal intervals. Consider, for example, the difficulty in developing items for a depression scale in which the intervals would always be equal throughout the scoring of the instrument. In counseling, many instruments are treated as if there were interval data. For example, the difference between 85 and 100 on an intelligence test is supposedly the same as the difference between 100 and 115. Many of the statistical techniques used to evaluate assessment instruments also assume that there is an interval scale. There are instruments, however, that do not have an interval scale and that should not be evaluated using those statistical techniques. In evaluating

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

instruments, counselors need to consider whether the scores can be classified on an interval scale because many of the statistical techniques discussed later in this book assume there are interval data. Sometimes a researcher will put together an instrument, for example, using a Likert scale, where it is very difficult to achieve interval data (e.g., strongly disagree, disagree, agree, strongly agree). In this example, is the difference between strongly disagree and disagree equal to the difference between disagree and agree? The researcher may then use statistics that require interval data when that is not appropriate. We should be somewhat cautious in using an instrument developed in this manner.

A **ratio scale** has all the properties of an interval scale, but the former requires the existence of a meaningful zero. A good example of a meaningful zero is driving and miles per hour. When you are driving at zero miles per hour, you are stopped, and not moving is a meaningful zero. With intelligence tests, there is no definition of what a score of zero means. Therefore, the results from an intelligence test might be considered an interval scale but not a ratio scale. In measuring other entities, such as pulse or liquid ounces, there are meaningful zeros. Weight is an example of a ratio scale because pounds are equal units and a meaningful zero exists. With this type of scale, it is possible to compute ratios and make conclusions that something is twice as big as something else. For example, a person who weighs 200 pounds is twice as large as someone who weighs 100 pounds. With interval scales, such as IQ, these types of conclusions cannot be made. Thus, we cannot say that someone whose IQ is 150 is twice as smart as someone whose IQ is 75.

As indicated earlier, the level of measurement influences the methods that can be used to evaluate the instrument. Therefore, counselors need to consider the scale of measurement when they begin to examine an instrument.

LO2 Norm-Referenced vs. Criterion-Referenced Instruments

To interpret a score on any instrument, practitioners first need to consider whether it is a norm-referenced or criterion-referenced instrument. In a **norm-referenced instrument**, an individual's score is compared with scores of other individuals who have taken the same instrument. The term *norm* refers to the group of individuals who took the instrument to which others' scores are then compared. A norming group can be quite large, such as a national sample of 2,000 adults who have taken a personality inventory. On the other hand, a norm could be your fellow classmates, such as when an instructor "grades on the curve." Simply stated, a norm-referenced instrument is one in which an individual's performance is compared with the performance of other individuals.

In a **criterion-referenced instrument**, the individual's score is compared with an established standard or criterion. Criterion-referenced instruments are sometimes called domain- or objective-referenced. With a criterion-referenced instrument, the interest is *not* on how the individual's performance compares with the performance of others, but rather on how the individual performs with respect to some standard or criterion. In criterion-referenced instruments, results are reported in terms of some specific domain. Therefore, to interpret a client's criterion-referenced test results, a counselor needs to understand the domain being measured, such as third-grade spelling words, multiplication of two-digit numbers, or knowledge of counseling theories. In criterion-referenced testing, the testing often pertains to whether a person has reached a certain *standard* of performance within that domain. For example, does Jerome get 70% of a sample of second-grade arithmetic problems correct? Does Lisa spell 90% of fifth-grade spelling words correctly? Many of the tests you have taken in your academic career have been

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

criterion-referenced tests on which, for example, you needed to score 90% or better for an A, 80% or better for a B, 70% or better for a C, and so forth.

Sometimes with a criterion-referenced test there is a *mastery* component. In these cases, a predetermined cutoff score indicates whether the person has attained an established level of mastery. For example, a state's department of education may designate a certain test score on a high school graduation examination that all high school students must achieve before they can graduate. Likewise, a teacher may test to find out if a student has mastered the multiplication table for the number 3 before allowing the student to learn the multiplication table for the number 4. Professional licensing examinations for counselors and psychologists are also examples of criterion-referenced tests that include a mastery component. To be licensed in many states, counselors must meet a preestablished score on a licensing examination that reflects a certain level of mastery. Individuals who do not attain that score are considered not to have mastered the body of knowledge necessary to perform as professional counselors.

There are several difficulties in developing sound criterion-referenced instruments. First, criterion-referenced tests are designed to measure a person's level of performance in a certain domain. Therefore, an instrument should adequately measure that domain. For example, to show mastery of a theories and techniques class in a counseling program, a criterion-referenced test for this class would need to adequately measure all the information in the domain of counseling theories and techniques. However, there is no universal agreement within the field about which theories are most important, nor is there complete agreement on which counseling techniques are most effective. Therefore, it is difficult to determine what content should be included on a theories and techniques test. With higher-order knowledge areas, it is often quite difficult to define definitively the precise content domain and determine the degree to which topics within the content domain are particularly or less important. Anastasi and Urbina (1997) contended that criterion-referenced testing is best for assessing basic skills, such as reading and mathematics at the elementary level.

Another problem with criterion-referenced instruments is determining the criterion for mastery level. This issue centers on what exact score truly indicates that someone has mastered the content. For example, if the mastery level is preset at 90%, does it, in effect, mean that a person who receives 89% has not mastered the content? What about the person who scores below the mastery level because of a mistake in marking an answer? It is often difficult to determine the exact score that indicates whether someone has, indeed, mastered that content. The consequences of making mistakes in misidentification can often influence the setting of the mastery level. For example, in a professional licensing examination (e.g., in counseling), the consequences of misidentification can be quite serious if individuals are considered, by their performance on the instrument, to be competent when in fact they are not adequately prepared to counsel clients. To keep this from happening, the cutoff score might be set quite high to minimize the number of false positives. Although there are often numerous difficulties in developing sound criterion-referenced instruments, these instruments can be very useful in indicating an individual's level of performance in a certain content domain.

Interpreting Norm-Referenced Instruments

Let's return once again to our example of the Counseling Aptitude Scale (CAS) and your score of 60. In this example, this fictitious instrument is used to identify which individuals have the greatest potential for being superb counselors as compared with other individuals. Because we are comparing the scores of individuals, it is a norm-referenced interpretation. Often with norm-referenced instruments, we use statistics to help

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 2.1
Raw scores on the Counseling Aptitude Scale

Names	CAS Scores	Names	CAS Scores
You	60	Tom	10
Alex	40	Angelina	50
Stacy	50	Max	40
Josh	50	Juan	50
Michael	70	Marilyn	20
Katlin	30	Satish	60
Morgan	60	Lisa	90
Ian	30	Omar	60
Christina	50	Matt	50
Hannah	80	Luisa	40
Marco	60	Rolando	90
Adam	50	Carmin	50
Rose	70	Yu-Ting	70
Kisha	50	Matt	50
Reggie	40	Logan	30
Kris	30	Regina	60
Bart	10	Colton	40
Rebecca	60	Punita	70
Scott	20	Javi	80
Shawn	60	Emily	40

interpret the scores. If we examine the scores in Table 2.1, it would be difficult to simply look at the scores and to interpret your score of 60 compared with the scores of 39 other people. However, organizing the scores in a logical manner can facilitate the process of understanding the scores.

L03 By converting the scores into a **frequency distribution**, as is reflected in Table 2.2, we can better understand how your score of 60 compares with the others who also took the fictitious CAS. A frequency distribution is simply where the scores (X) are indicated on the first line and the frequency (f), or number, of people achieving that score is indicated beneath. A frequency distribution provides a visual display to help organize data so that it is easier to see how the scores are distributed. In the frequency distribution in Table 2.2, we can see that the number, or frequency, of people receiving scores of 10, 20, 80, and 90 is smaller than the number of people receiving scores of 40, 50, and 60.

Sometimes, actually graphing the frequency of scores can provide a better visual display of how a group of people scored on an instrument. A **frequency polygon** is often used in assessment because this graphic representation makes the data easier to understand. A frequency polygon is a graph that charts the scores on the x-axis (the horizontal) and the frequencies of the scores on the y-axis (the vertical). A point is placed by plotting the number of persons receiving each score across from the appropriate frequency. The successive points

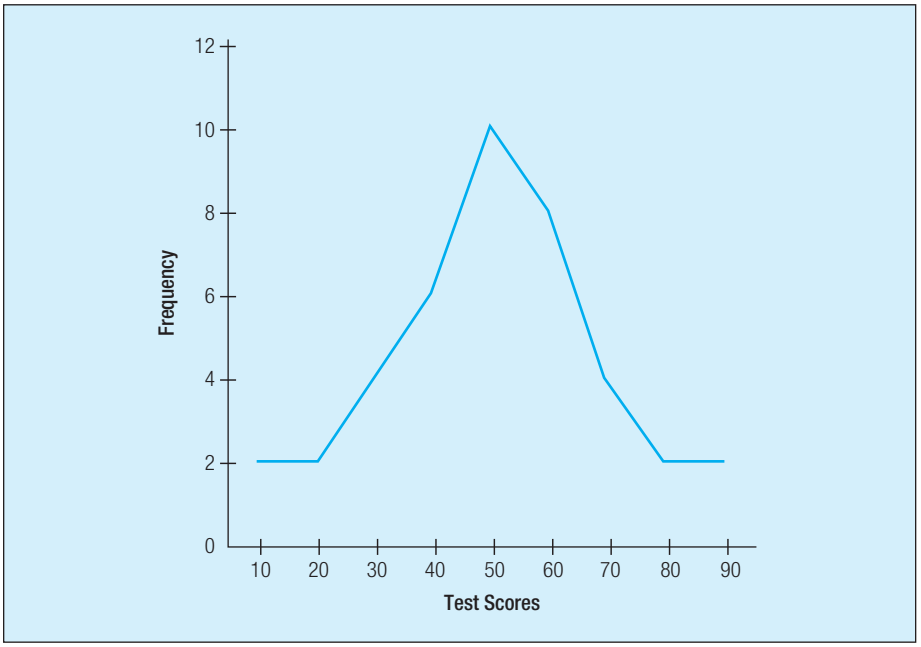
TABLE 2.2
Frequency distribution for the scores on the Counseling Aptitude Scale

X	10	20	30	40	50	60	70	80	90
f	2	2	4	6	10	8	4	2	2

Note: X is for CAS scores, and *f* is for the frequency of individuals.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 2.1
Frequency polygon for Counseling Aptitude Scale scores



are then connected with a line. Figure 2.1 shows a frequency polygon for the CAS scores and reflects a graphic display of the information contained in the frequency distribution in Table 2.2. The frequency polygon in Figure 2.1 makes it easier to see that many more people had a score lower than your 60, although some individuals did have scores higher than yours. This frequency polygon also reflects that most of the people scored between 40 and 60, with the largest number of people receiving a score of 50.

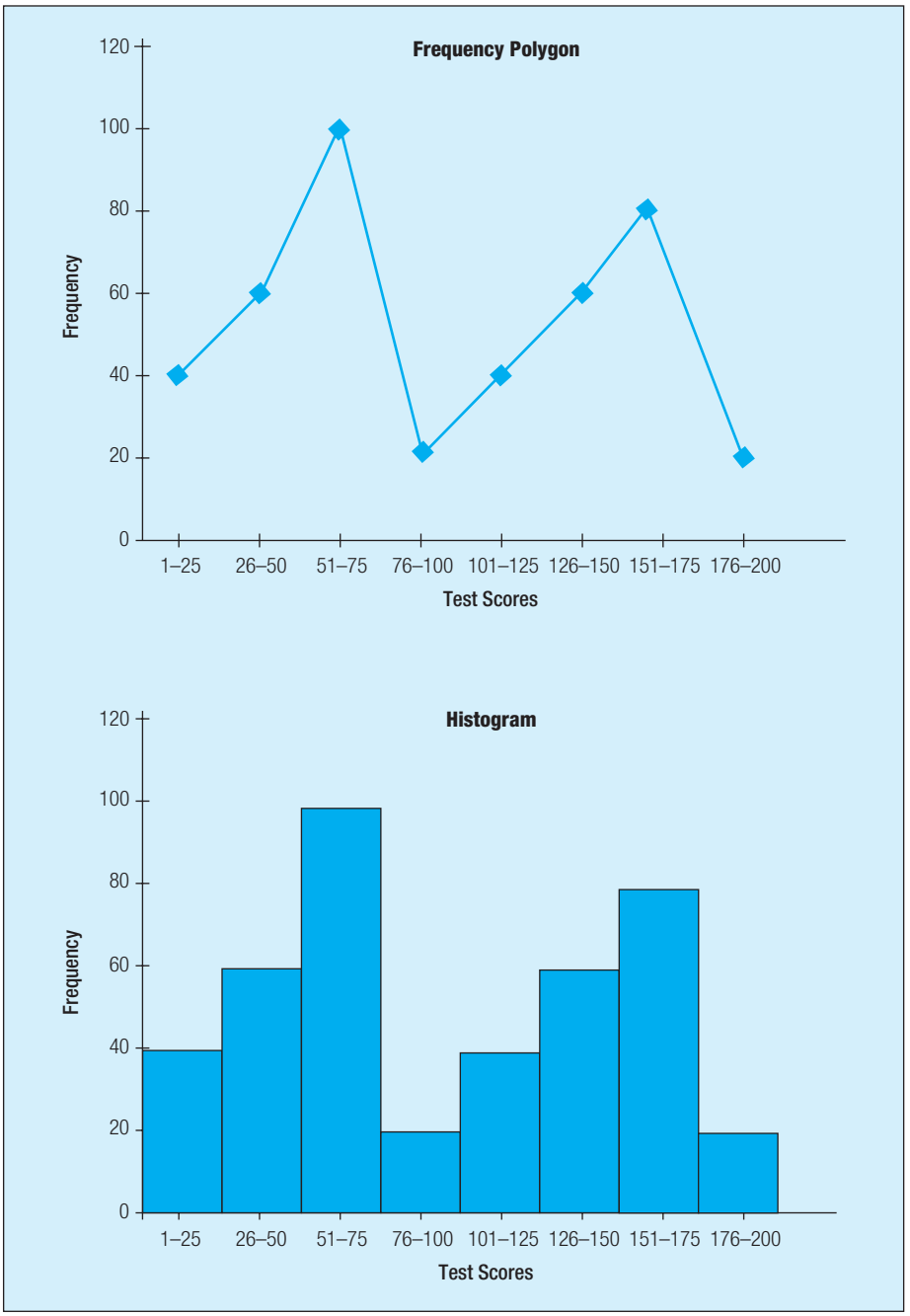
Sometimes, we do not want to plot each individual score of an instrument because the range of scores is too large. For example, an instrument with a range of scores from 1 to 200 would make a large frequency polygon if we plotted each possible score (1, 2, 3, … , 198, 199, 200). In this case, we may want to make the information more manageable by determining the frequency of people who fall within a certain interval of scores. Figure 2.2 contains a frequency polygon that groups the scores into convenient interval classes of 25 points each (e.g., 1–25, 26–50, 51–75). In a **histogram** (see Figure 2.2), the height of the column over each class interval corresponds to the number of individuals who scored in that interval. As Figure 2.2 indicates, either a frequency polygon or a histogram can be used to provide a visual display of the same frequency of scores. The choice of which method to use depends on which provides the best graphic display.

L04 Measures of Central Tendency

Graphing the scores by using a frequency polygon or a histogram can help you understand how your score of 60 on the Counseling Aptitude Scale compares with the scores of others. It also may be useful to have some indication of how *most* of the other people performed on the CAS. The mode, median, and mean are measures of central tendency that can be useful in interpreting individuals' results on an instrument. The **mode** is the *most frequent* score in a distribution. On the CAS scores, the mode, or the most frequent score, is 50, because both Tables 2.1 and 2.2 reflect that 10 individuals received this score, which is the highest frequency of any of the scores. To find the mode, you merely count the number of people who received each score, and the score with the highest number of people is the mode.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 2.2
Frequency polygon and histogram with interval scores



You now know that your score of 60 is 10 points higher than the mode. It might also be helpful to know if your score is above or below the median. The **median** is the score at which 50% of the people had a score below it and 50% of the people had a score above it. In other words, the median is the score that evenly divides the scores into two halves. You determine the median by arranging the scores in order from lowest to highest and finding the middle number. If the distribution is comprised of an odd number of participants, then it is easy because you simply arrange the scores from smallest to largest and find the middle

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

number. For example, if there are 25 score in a distribution, then you would arrange the scores in order and then count to the 13th score. The 13th score would be the median because 12 people (50%) would have a lower score and 12 people (50%) would have a higher score. In the CAS example, there are 40 scores and thus there will not be one middle number. In this case, you take the average between the two middle scores. In the CAS examples, this is easy because when the scores are arranged from smallest to largest, the two middle scores are 50 (and $50 + 50 \div 2 = 50$) and, therefore, the median is 50. As another example, let's say we have the following scores: 1, 1, 1, 3, 14, 16. The median would be 2 because that is the average of 1 and 3, which are the two middle scores in this distribution.

Although a mode of 1 and a median of 2 provide some useful information about the scores of 1, 1, 1, 3, 14, 16, these measures of central tendency do not really describe the variation in scores very well, because the higher scores of 14 and 16 are not really reflected in either the mode or the median. This is an example of why the mean is a useful measure of central tendency. The **mean** is the arithmetic average of the scores. The mean in the example of 1, 1, 1, 3, 14, 16 is 6, which provides a measure that represents the distribution of scores. In this book, which includes a few statistical formulas, the mean will always be notated by the letter M and an individual's score or some other data point by X. The Greek symbol Σ stands for sigma and means "to sum," and N represents the number of individuals in the group. In the formula for mean (M), we sum (Σ) the scores (X) and then divide by the number of individuals in the group (N) to get the mean or average (M). In the earlier example, we would sum the scores ($1 + 1 + 1 + 3 + 14 + 16 = 36$) and then divide 36 by the number of scores ($N = 6$), which is equal to 6. In case you are interested, the mean for the CAS scores listed in Table 2.1 is 50.50, whereas the median and the mode are both 50.

$$M = \frac{\Sigma X}{N}$$

LO5 Measures of Variability

Knowing the measures of central tendency assists you in understanding what your score of 60 on the Counseling Aptitude Scale means, because you now know that your score is larger than the mean, median, and mode. At this point, though, you cannot tell how much above the mean your score is in comparison with other people's scores. For example, it would suggest something very different if the 25 other people who took the Counseling Aptitude Scale all had scores of 60 as compared with the distribution of scores that was presented earlier (see Table 2.2). In assessments, it is important to examine how the scores vary so that we can determine if a person is high or low compared with others, and how much higher or lower a person's score is. For example, the scores of 1, 3, 6, 9, and 11 have a mean of 6, and the scores of 5, 6, 6, 6, and 7 also have a mean of 6. Even though the means are the same, the variations in scores affect how we would interpret a score of 7. Measures of variability provide useful benchmarks because they give us an indication of how the scores vary.

Range

Range provides a measure of the spread of scores and indicates the variability between the highest and the lowest scores. Range is calculated by simply subtracting the lowest score from the highest. On the Counseling Aptitude Scale, the highest score anyone received was 90, and the lowest was 10; therefore, the range is 80 ($90 - 10 = 80$). Range does provide an indication of how compact or wide the variation is, but it does not really assist us in interpreting your CAS score of 60 very well. Range is a simple and somewhat crude measure that can be significantly influenced by one extremely high or one very low score.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Variance and Standard Deviation

When interpreting scores to clients, counselors often use more precise measures to provide better information about how individuals vary from the mean. Variance and standard deviation fill this need by providing more precise measures that serve as indicators of how scores vary from the mean. In examining Table 2.3, we might be interested to know how these five scores (1, 2, 3, 4, 5), on average, vary from the mean. In the second column, there is an indication of how each score varies from the mean of 3. If we wanted an average deviation from the mean, we could simply add these numbers together and divide by the number of scores. If we do that, however, what we get is 0. In fact, with any set of scores, given the way we calculate mean, we will always get 0. Therefore, to avoid this problem, we square the deviations, add these together, and divide by the number of scores. This number is **variance** or mean square deviation. In the case of Table 2.3, by adding the squared deviation in the third column and dividing the number of scores, the variance is 2. The problem with variance is that when we square the deviations, they are no longer in the same measurement unit as the original scores. Therefore, if we take the square root of the variance, it will take the deviation back to the original unit of measurement. The square root of variance is the **standard deviation**, which provides an indication of the average deviation from the mean in the original unit of measurement. In Table 2.3, the square root of 2 (variance) is 1.41, which is the standard deviation of those scores.

In the area of assessment, standard deviation is primarily used in two ways. First, standard deviation provides some indication of the variability of scores. The larger the standard deviation, the more the scores varied, or deviated, from the mean. Smaller standard deviations indicate that the scores are more closely grouped together around the mean. Based on the scores presented earlier, the standard deviation for the CAS is 19.10. A counselor cannot ascertain how much variation there is in scores based solely on the standard deviation; however, it is probably safe to say that this standard deviation of 19.10 suggests that there is some variation in scores on the CAS.

The second way to use standard deviation involves interpreting individual scores. A standard deviation can provide an indication of whether a score is below the mean, close to the mean, or significantly higher than the mean. We can interpret your score of 60 on the CAS by saying that your score is quite close to one half a standard deviation above the mean (60 − 50.50 = 9.50), because one half a standard deviation for the CAS is 9.55 (19.10/2 = 9.55).

TABLE 2.3
Illustrations of variance and standard deviation

Score (X)	Deviation ($X - M$)	Deviation^2 ($X - M$)^2
1	$1 - 3 = -2$	$-2^2 = 4$
2	$2 - 3 = -1$	$-1^2 = 1$
3	$3 - 3 = 0$	$0^2 = 0$
4	$4 - 3 = 1$	$1^2 = 1$
5	$5 - 3 = 2$	$2^2 = 4$
$\Sigma X = 15$	$\Sigma(X = M) = 0$	$\Sigma(X = M)^2 = 10$
$M = \frac{\Sigma X}{N}$		$S^2 = \frac{\Sigma(X - M)^2}{N}$ $S^2 = \frac{10}{5} = 2$
$M = \frac{15}{5} = 3$		$S = \sqrt{\frac{\Sigma(X - M)^2}{N}}$ $S = \sqrt{2} = 1.41$

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

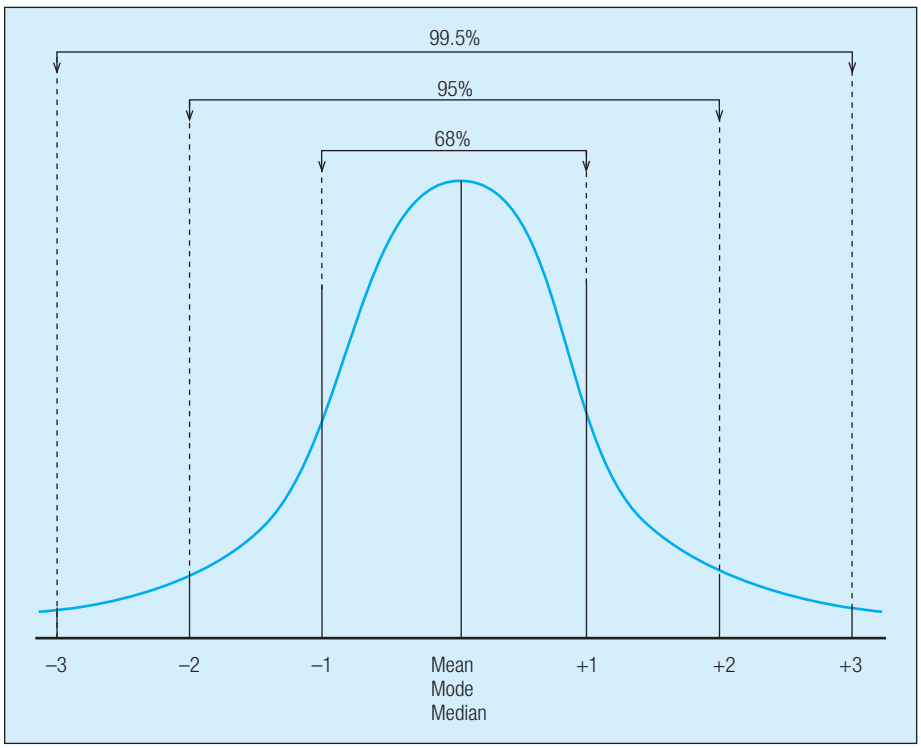
L06 Normal Distributions

If the scores on an instrument fall into a **normal distribution**, or a **normal curve**, then standard deviation provides even more interpretive information. The distributions of a number of human traits (e.g., weight, height, intelligence, and some personality characteristics) approximate the normal curve. The normal curve has mathematical properties that make it very useful in the interpretation of norm-referenced instruments.

The normal curve is bell-shaped and symmetrical (see Figure 2.3). If the distribution of scores is normalized, there is a single peak in the center and the mean, mode, and median are the same score. The normal curve reflects that the largest number of cases falls in the center range and that the number of cases decreases gradually in both directions. Specifically, 68% of the cases fall between 1 standard deviation below the mean and 1 standard deviation above the mean. Ninety-five percent of the cases fall between 2 standard deviations below the mean and 2 standard deviations above the mean, and 99.5% of the cases fall between 3 standard deviations below the mean and 3 standard deviations above the mean. The percentages of people falling within the different ranges of standard deviations are constant and have a normal distribution. In general, the larger the group, the greater the likelihood that the scores on an instrument will resemble a normal curve. If a counselor is using an assessment where the scores have a normal distribution, then a counselor interprets individuals' scores based on the characteristics of a normal distribution. For example, if an individual score is more than 3 standard deviations above the mean, we know that only a very small percentage of people score there, because 99.5% of the individuals score between 3 standard deviations below the mean and 3 standard deviations above the mean.

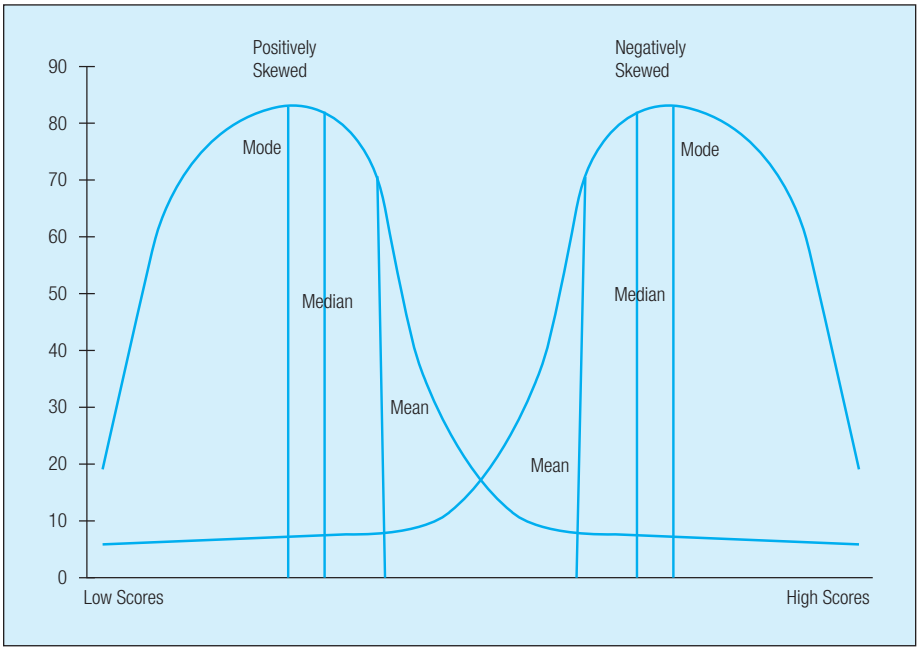
There are times, however, when the scores do not fall into a normal distribution, so using the normal curve to interpret scores would be inappropriate. Scores on an instrument can distribute in many ways other than a normal curve. They can be multimodal,

FIGURE 2.3
A normal distribution



Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 2.4
Skewed distributions



which means the distribution has more than one peak. Distributions can also be **skewed** (see Figure 2.4), meaning that the distribution is not symmetrical and the majority of people either scored in the low range or the high range, as compared with a normal distribution in which the majority scored in the middle. A **positively skewed distribution** is one in which the majority of scores are at the lower end of the range of scores. A **negatively skewed distribution** is the opposite, where the majority of scores are on the higher end of the distribution. Some students in an assessment course memorize the difference between positively and negatively skewed distributions simply by thinking that it is the opposite of what makes sense. You may use this technique for learning the difference, but there is a better explanation. If you look at Figure 2.4, you will notice that the mean is more positive as compared with the mode and median in the positively skewed distribution. The opposite occurs with the negatively skewed distribution, where the mean is more negative than either the mode or the median. Thus, a distribution is labeled either positive or negative, because the mean is skewed positively when the scores are mostly lower, and the mean is skewed negatively when the scores are mostly higher. Determining if the scores on an instrument are skewed is important, because professionals will sometimes interpret scores assuming there is a normal distribution when the distribution is actually skewed.

Types of Scores

Counselors often need to provide clients with interpretations of assessment results. For example, a school counselor interprets achievement test scores to students or parents, a mental health counselor interprets a personality inventory to a client, or a marriage and family counselor interprets a family functioning inventory to a family. Sometimes the manner in which test results are communicated to individuals can be quite confusing because clients have little exposure to terms such as *normal distribution*, *percentile*, or *stanine*. Therefore, counselors need to understand the different methods of scoring and become adept at communicating those results in ways that clients can understand.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 2.4
Frequency distribution for the scores on the Counseling Aptitude Scale

X	10	20	30	40	50	60	70	80	90
F	2	2	4	6	10	8	4	2	2
%	5	5	10	15	25	20	10	5	5
Percentile	5	10	20	35	60	80	90	95	99*

*The highest percentile is the 99th percentile.

The simplest of the scoring methods is the **raw score**, which is meaningless without any other interpretative data. We get raw scores by simply following the rules for scoring. Raw scores are those scores that have not been converted to another type of scoring (e.g., percentile, *T* score). As indicated earlier in this chapter, your raw score of 60 on the Counseling Aptitude Scale actually means nothing until we began to examine the distribution of scores, measures of central tendency, and measures of variability. When we examine these indicators, we typically have more information with which to interpret a score.

LO7 Sometimes, however, we want additional information concerning how an individual performed in comparison with the norming group, and this often is accomplished by converting the raw score to a percentile rank or percentile score. **Percentile scores**, or **percentile ranks**, indicate the percentage of people in the norming group who had a score *at* or *below* a given raw score. Percentiles should not be confused with the common scoring method of indicating the percentage of items answered correctly. Percentiles can be determined for any distribution of scores, not just for normalized distributions. Table 2.4 reflects how percentiles can be calculated for the CAS scores. As you probably noticed, the first two lines of Table 2.4 are the same as the frequency distribution in Table 2.2. To find the percentile, we first have to determine how many people out of the group had each score, and then we calculate the percentage of people out of the total group who received that score. For example, for the score of 50, we can see that 10 people received that score, and 10 divided by the total group of 40 is equal to 25%.

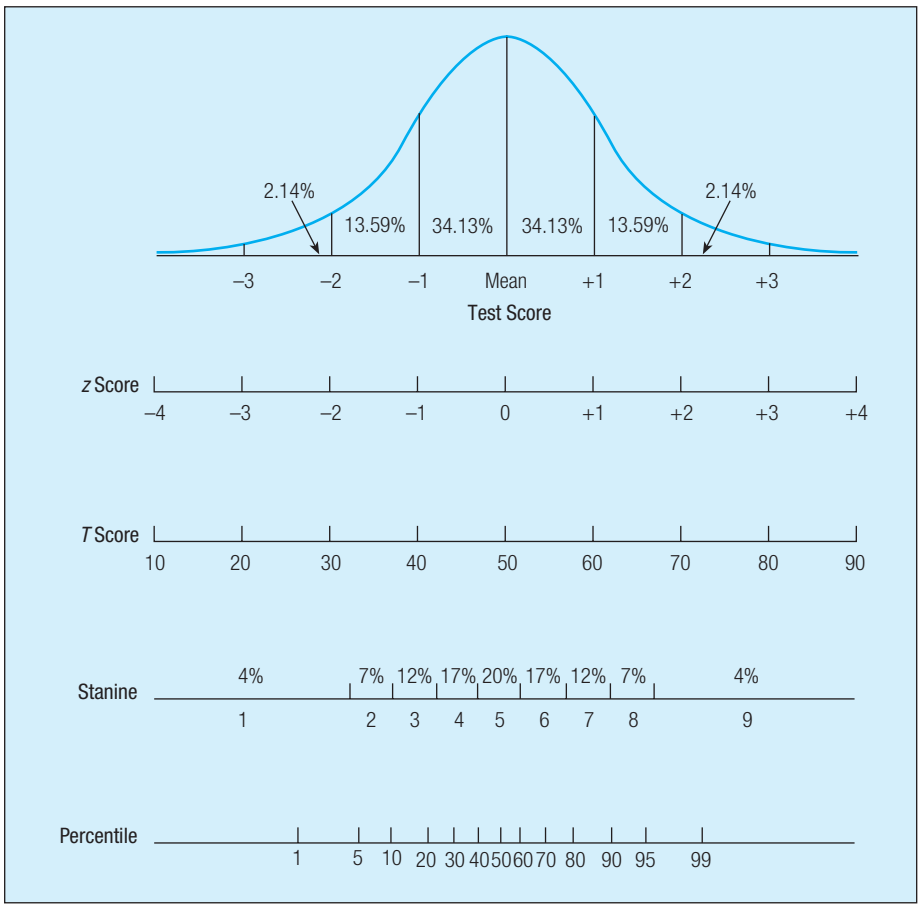
As stated earlier, percentiles are the *percentage* of people who received a score *at* or *below* a given raw score. Therefore, to get percentiles, we add the *percentages* of individuals who had a score *at* or *below* each of the raw scores. If we examine Table 2.4, we can see that the percentile for a score of 10 is the 5th percentile, because 5% of the people received this score (2/40 = .05) and no people had a score lower than this. We would interpret this score to a client by saying, "If a hundred people had taken this instrument, five people would have a score at or below yours." Let's look at your score (60) and see what your percentile would be. The percentage of people who received each of the scores is listed underneath the frequencies. Thus, to determine the percentile, we merely add the percentage of people at each score that are at or lower than 60 (5 + 5 + 10 + 15 + 25 + 20 = 80). In interpreting your score of 60, a counselor might say, "Eighty percent of the norming group had a score at or below yours." Another way of communicating this result is to say, "If 100 people took the CAS, 80 of them would have a score at or below yours and 20 would have a higher score." In many ways, using percentiles may help you understand your score of 60 better than some of the other methods we have discussed so far.

Percentiles are used in many commonly used counseling instruments that have a normal distribution. Sometimes on such instruments as the Armed Services Vocational Aptitude Battery (ASVAB), the results are presented using a percentile band that is based on standard error of measurement (see Chapter 3). Percentile bands are useful because they can show a range of where clients could expect their scores to fall if they took the instrument multiple times.

Although percentiles are easy to understand and provide some useful information, there are drawbacks. Percentiles are not in equal units, particularly at the extremes of the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 2.5
The normal distributions and various types of standardized scores



distribution if the raw scores approximate a normal curve. Looking at the bottom of Figure 2.5, you will notice that there is a large difference between the 1st and 5th percentiles, whereas there is a very small difference between the 40th and 50th percentiles. With a normal distribution, scores will cluster near the median and mean, and there will be very few scores near the ends of the distribution. An example of the lack of equality in units can be seen from the distribution of CAS scores. As you will notice in Table 2.4, the difference in percentile scores between the raw scores of 80 and 90 is 4 (99 − 95 = 4). On the other hand, the difference in percentile scores between the raw scores of 40 and 50 is 25 (60 − 35 = 25). Consequently, the same 10-point difference in raw scores results in remarkable differences in percentile scores. Thus, percentile scores are helpful in providing information about each individual relative position in the normative sample, but they are not particularly useful in indicating the amount of difference between scores. In addition, percentiles are an ordinal scale and cannot be analyzed using statistics that require interval data.

LO8 Standard Scores

Standard scores address the limitation of unequal units of percentiles, while providing a "shorthand" method for understanding test results. Standard scores are different methods of converting raw scores so that there is always a set standard deviation and mean. Standard scores can be used with all types of instruments, including intelligence, personality, and career assessments. As a result, when we hear that a client has a z score of +1.00 or a

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 2.5
Calculations of z scores

z Score Formula	Sample Text
$Z = \frac{X - M}{s}$	$M = 75$
	$S = 5$
Abby's Score	**Keith's Score**
$X = 85$	$X = 72$
$Z = \frac{85 - 75}{5} = 2.00$	$Z = \frac{72 - 75}{5} = -.60$

T score of 60, we, as professional counselors, will automatically know where that score falls within the distribution without needing more detailed information (i.e., mean and standard deviation) on that specific instrument. Standard scores, therefore, provide a quick, easy way to know a client's relative position on an instrument because standard scores describe how many standard deviations a client's score is from the mean.

z Scores

There are various standard scores (e.g., *T* scores, deviation IQs), but all of them are based on **z scores**. To convert an individual raw score into a *z* score, we subtract the mean of the instrument from the client's raw score and divide by the standard deviation of the instrument. The formula for computing a *z* score is:

$$z = \frac{X - M}{S}$$

where *z* is the standard score, *X* is an individual's raw score, *M* is the mean of the instrument or test, and *s* is the standard deviation of that instrument. Because *z* scores are standard scores, there is always a set mean and a set standard deviation. With *z* scores, the mean is always 0 and the standard deviation is always 1. Table 2.5 illustrates the calculation of *z* scores for two individuals on the same instrument that has a mean of 75 and a standard deviation of 5. Abby's *z* score of 2.00 indicates that her score is 2 standard deviations *above* the mean, and Keith's *z* score of −.60 reflects that he is a little more than one half a standard deviation *below* the mean. With *z* scores, it is important for counselors to focus on whether the *z* score is positive or negative. A counselor would not want to communicate to a client that his or her score was above the mean when in actuality it was below the mean or vice versa.

If an instrument has a normal distribution, then *z* scores are considered to be normalized and can provide additional information. Therefore, counselors can then visualize the normal curve and identify where a particular *z* score would fall. For example, we can see that Abby's *z* score of 2.00 is 2 standard deviations above the mean, and we can see that a large percentage of people, 97.72% to be exact, would have a score at or below Abby's. Keith's *z* score of −.60 is below the mean of 0, and a smaller number of people (27.43%) have a score at or below his. We can find the exact percentile score for each *z* score by using the conversion table in Appendix B, which lists the percentage of people who would fall at or below each *z* score.

T Scores

T scores, another standard score, have a fixed mean of 50 and a standard deviation of 10. A *z* score can be converted to a *T* score by multiplying the *z* score by 10 and adding or subtracting that number from 50. The *z* score is considered the base of standard scores

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

because *z* scores are typically used when raw scores are converted to other standard scores. For example, some test developers prefer to use *T* scores rather than *z* scores because they wanted to eliminate the decimals and positive and negative signs of *z* scores.

Normalized *T* scores have the same advantages as normalized *z* scores (see Figure 2.5). Although *T* scores are often normalized standard scores, it is important for counselors to check the manual of an instrument to ensure that this assumption is correct.

Some instrument developers will transform the data to normalize the distribution, which leads to an interesting dilemma concerning when it is legitimate to transform a nonnormal distribution into a normal distribution and when it is unacceptable to normalize a distribution. Anastasi and Urbina (1997) suggested that transforming a nonnormal distribution should be done only with very large samples and when the lack of normality is related to modest limitations of the instrument. Questions have been raised about some instruments and the transforming of the distribution when the construct is not normalized. As an example, some researchers have questioned whether marital satisfaction in the population really distributes normally; that is, are the levels of satisfaction for most couples around the mean, with very few people being very unhappy or very happy? Some instruments, such as the Marital Satisfaction Inventory—Revised, have been criticized for using transformed *T* scores on constructs that may not be normalized in the population.

Stanines

Another normalized standard score is the stanine scale; *stanine* is a contraction of *stan*dard and *nine*. Stanines, which range from 1 to 9 with a mean of 5 and a standard deviation of approximately 2, are different from other standard scores because they represent a range of percentile scores. Table 2.6 indicates the percentile scores and the percentage of people who fall within each stanine score. Raw scores are converted to stanines by having the lowest 4% of the individuals receive a stanine score of 1, the next 7% receive a stanine of 2, the next 12% receive a stanine of 3, and so on, through the group.

A stanine has the advantage of being a single-digit number that many people find easy to use. The disadvantage is that stanines represent a range of scores, and sometimes people do not understand that one number represents various raw scores. For example, if Josh scored at the 40th percentile and Maria scored at the 41st percentile, Josh's stanine score would be 4 and Maria's would be 5. On the other hand, if Sam scored at the 24th percentile, he would get the same score (i.e., 4th stanine) as Josh. Therefore, when information is needed for very important decisions in a client's life, practitioners usually should employ more precise measures than stanines.

TABLE 2.6
Stanines

Stanine	Percentile Range	% of Individuals
1	1 to 4	4
2	5 to 11	7
3	12 to 23	12
4	24 to 40	17
5	41 to 59	20
6	60 to 76	17
7	77 to 88	12
8	89 to 95	7
9	96 to 99	4

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Other Standard Scores

There are other standard scores, such as deviation IQs and College Entrance Examination Board (CEEB) scores, that mental health practitioners use. Deviation IQs are an extension of the ratio IQ, which was used in early intelligence tests. The earlier quotients were a ratio of mental age to chronological age times 100 (IQ = MA/CA × 100). If an individual's mental age were the same as her chronological age, her IQ score would be 100, a number that was considered the benchmark for normal performance. One of the difficulties with the ratio method of calculating IQ, however, is that it does not produce interval data. Furthermore, because different age groups have different standard deviations, it is difficult to remember all the different standard deviations. Therefore, deviation IQs have now replaced ratio IQs in intelligence tests. Deviation IQs are standard scores in which the deviations from the mean are converted into standard scores, which typically have a mean of 100 and a standard deviation of 15.

Another commonly used standard score is the CEEB score. In this standard score, the raw scores are converted into standard scores with a mean of 500 and a standard deviation of 100. Many people are familiar with this standard score, although they may not be familiar with the title "CEEB score," because instruments such as the SAT use this standard score.

Age or Grade Equivalent Scores. There are two common methods for calculating **age or grade equivalent scores**. One is through a procedure called **item response theory** (see Chapter 4) in which items are specifically developed to measure performance at certain developmental levels. The second method of calculating age norms or age equivalent norms is with a norm-referenced approach in which the focus is on comparing individuals with other individuals. Often, rather than comparing the individual with other age level performances, the process simply involves comparing the individuals with other individuals at that same age. The average performance score of 9-year-old children becomes the equivalent of a score of 9. Different age level equivalents are extrapolated depending on whether the individual is above or below the mean. Therefore, when this process is used to determine an age equivalent score, we should not interpret that a score of 12 reported for a 9-year-old student means that the child is functioning as an average 12-year-old. Rather, the 9-year-old student is being compared with other 9-year-old children and is performing at a considerably higher level than the average 9-year-old.

Grade equivalent norms, or grade equivalent scores, are based on the same reasoning as age norms. With grade norms, however, the average score for students in a particular grade is converted into the grade equivalent for that specific grade. Grade equivalent scores are widely used in the area of achievement testing, primarily because they are easy to communicate. It is often easier to say a child is performing at the sixth grade, second month (6.2) level than at the 79th percentile. Parents and others like grade equivalent scores because there is more of a context for interpreting those scores. For example, at a party, a parent can simply brag that little Jimmy is already reading at the fourth-grade level, and he is only in first grade. The problem with this is that Jimmy is probably not reading at the fourth-grade level; rather, his score indicates that he performed substantially higher than the average first grader. Moreover, a reading test for first graders may not even include any items that are truly fourth-grade reading items. Grade equivalent scores are typically extrapolated scores based on average performance of students at that grade level. Therefore, Jimmy's grade equivalent score does not indicate that Jimmy should be placed in a fourth-grade class for reading.

Professionals in the counseling field need to be aware of other problems with age norms and grade norms. A significant problem with extrapolated age and grade equivalent

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

scores is that some professionals believe these scores reflect precise performance on an assessment, but that is not actually accurate. As a result of various problems with age or grade equivalent scores, Cronbach (1990) condemned the use of them.

The information that age and grade equivalent scores communicate is often ambiguous because there is no universal agreement on what constitutes age or grade achievement. For example, does 11th-grade mathematics include trigonometry or a combination of algebra and geometry? Thus, one person's interpretation of a grade equivalent could be very different from another's interpretation.

A second problem with age and grade norms is that learning does not always occur in nice, equal developmental levels. Let's take the example of reading. Consider whether the difference in reading that typically occurs between first and second grades is equal to the change that takes place between seventh and eighth grades. Many people would contend that the change that takes place between first and second grades is far greater than the change between seventh and eighth grades; therefore, we do not have equal units of measurement in age or grade norms.

Another problem with age and grade equivalent scores is that instruments will vary in the scoring. One publisher's test could give a child a sixth grade, eighth month score (6.8), and another publisher's instrument could result in a score of 7.1. Although the two scores may be related to small differences between the instruments, consumers of the scores may have very different interpretations of scores that are really not all that discrepant. Another problem with age or grade equivalent scores is that teachers or administrators may expect all students to perform at or above their respective age or grade level. For example, teachers have been reprimanded because students have had scores below grade level.

These misconceptions fail to take into account that the instruments are norm-referenced; thus, the expectations are that 50% of the students will fall above the appropriate age or grade score and 50% will fall below this score. Therefore, in most classrooms, expecting all students to fall above the mean is unrealistic as well as inappropriate given norm-referenced testing.

Because age and grade equivalent scores are often misunderstood, counselors can be particularly helpful to parents or guardians by assisting them in understanding these scores. A case in point is Justin, a 10-year-old with academic and behavioral problems in school. Justin's parents entered counseling because they were having a major disagreement about whether Justin should repeat the fourth grade. At the beginning of the fourth grade, Justin's test scores showed that he had a reading grade equivalent of 3.4 (third grade, fourth month) and a language arts grade equivalent of 3.8. The father therefore anticipated that the counselor would automatically agree with him concerning the need for Justin to repeat the grade. The counselor explained to the parents that Justin was only slightly behind the average performance of fourth graders in language arts and a little more behind in reading. After working with the family for a few sessions and talking to Justin's teacher, the counselor realized that a number of complex issues were influencing Justin's academic and behavioral problems. Through counseling, the parents began to look at their parenting styles and found different methods they could use to assist Justin with his schoolwork. Although this case is an example of a situation in which the score is below grade level, parents also react to scores above grade level and want their child to be placed in a higher grade based on their child's grade equivalent test scores. Therefore, it is important for counselors to examine the instrument's manual and see how the age or grade equivalent scores are calculated. If the instrument developers used item response theory, then typically the counselors can have more confidence in the age or grade equivalent scores. However, if the instrument developers used the extrapolation method, then the counselor should try and avoid using age or grade equivalent scores.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L09 Evaluating the Norming Group

When using a norm-referenced instrument, it is important not only to consider what type of scoring is most appropriate (e.g., percentiles vs. *T* scores) but also to determine whether the norming group is suitable for the clients. You may wonder what the best type of norming group is; however, the answer to this question is dependent on a number of factors. The adequacy of a norming group depends on the clients being assessed, the purpose of the assessment, and the way in which that information is going to be used. Hence, there are *no* universal standards for what constitutes a good norming group; rather, the determination of an adequate norming group rests with the practitioner using the instrument. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) clearly indicates that the norming group should be adequately described so that test users can evaluate whether it is appropriate for their uses.

One way to evaluate a norming group is to examine the methods used for selecting the group. In general, more credence is given to a sample that is drawn in a systematic method. In statistics, the word *sample* is often used to refer to the group that is actually being tested. This sample is drawn from the larger group of interest, or the *population*. Returning to the Counseling Aptitude Scale example, the population would be the large group of all counseling students in the United States. It would be very difficult to have a norming group that included all of these counseling students; therefore, we might draw a sample of these students to represent this larger population.

One method of selecting a sample is to draw a **simple random sample** in which every individual in the population has an equal chance of being selected. Hence, if we wanted to use simple random sampling procedures with the CAS, we would have to give every counseling student in the United States an equal chance of being selected. Ensuring that every counseling student is entered into the pool from which the sample is drawn would be an extraordinarily difficult task.

A **stratified sample** is often used in the area of assessment. In a stratified sample, individuals are selected for the norming group based on certain demographic characteristics. For example, instrument developers may want their norming group to match the percentage of African Americans, Latino(a)s, Native Americans, and Asian Americans in the nation. They would need to actively recruit a sufficient number of people for their norming group so that the proportions of individuals from these ethnic groups match the proportions found in the latest reports from the U.S. Census Bureau. Other demographic variables that commonly influence the gathering of stratified samples include gender, socioeconomic level, geographic location, amount of education, marital status, and religion. Sometimes an instrument will incorporate a *stratified sampling technique* for a specific purpose, such as selecting children with certain learning disabilities or including a specified proportion of both depressed and nondepressed adults. If a stratified sample is used, the counselor should evaluate not only the adequacy of the sample but also the methods used for recruiting the sample. For instance, there are some instruments for adults in which the developers have recruited primarily college students rather than a wider sample of adults.

Another sampling strategy commonly used in the assessment field is **cluster sampling**, which involves using existing units or clusters rather than selecting individuals. By way of illustration, let's assume we are developing a measure of social skills for elementary age children. It would be unusually arduous to attempt a random sample and list all the elementary children even in one state. Therefore, the instrument developers might use a cluster sampling technique, where from a listing of all the elementary schools in a state, the instrument developer would randomly select a sample of elementary schools. The specific elementary students would not be randomly sampled; instead, larger units (i.e., elementary schools) would be randomly selected through cluster sampling.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

When a counselor uses a norm-referenced instrument, the interpretation of the results is restricted based on the particular normative population from which the norm was derived. The size of the norming group should be sufficiently large to provide a solid comparison. To use an instrument appropriately, counselors need to know about the characteristics of the norming group. An instrument's manual should include precise descriptions of the population that was sampled. Most manuals now include detailed information on characteristics of the norming group, particularly in relation to gender, race or ethnicity, educational background, and socioeconomic status. Test developers are responsible for informing the test user if there are situations in which the norms are less appropriate for some groups or individuals than for others (AERA, APA, & NCME, 2014). The counselor or test user, however, has the ultimate responsibility for determining whether the norming group is appropriate for clients. If there are discrepancies between the norming group and the client, the counselor needs to carefully interpret the results and inform others of the need to use that information cautiously.

Summary

This chapter summarizes some of the basic methods by which instruments are categorized, scored, and interpreted. In examining an instrument, it is important to determine if it is norm-referenced or criterion-referenced. Norm-referenced instruments compare an individual's performance with other people's performance on the instrument. Alternatively, criterion-referenced instruments compare an individual's performance with an established standard or criterion. In regard to norm-referenced instruments, statistics assist us in organizing and interpreting the results. Frequency distributions are charts that summarize the frequencies of scores on an instrument. Frequency polygons and histograms graphically display the distribution of scores so that practitioners can easily identify trends. Measures of central tendency (i.e., mode, median, and mean) provide benchmarks of the middle or central scores. The mean or average score is often used to interpret assessment results.

The measures of variability (i.e., range, variance, and standard deviation) indicate how scores vary and where an individual's score may fall in relation to the scores of others. Standard deviation is the most widely used measure of variability in assessment. Standard deviation has some important qualities if the distribution approximates the normal curve. With a normal distribution, 68% of the norming group will fall between 1 standard deviation above the mean and 1 standard deviation below the mean. With a normal distribution, the mean, mode, and median will all fall at the same point.

There are numerous methods for transforming raw scores in norm-referenced instruments. Percentiles provide an indication of what percentage of the norming group had a score at or below a client's. Standard scores, which convert raw scores so that there is always a set mean and standard deviation, express an individual's distance from the mean in terms of the standard deviation of the distribution. The most basic standard score is the *z* score, which has a mean of 0 and a standard deviation of 1. Counselors need to be careful when they use instruments that incorporate age equivalent or grade equivalent norms; they must know precisely how the results were calculated and how to interpret them appropriately. Furthermore, counselors need to fully examine the norming group of any instrument and understand the strengths and limitations of that group.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 3

Reliability

Learning Objectives

- **LO1** Describe what a reliability coefficient means using classical test theory
- **LO2** Understand the statistical technique of correlation
- **LO3** Explain the advantages and limitation of the major types of reliability coefficients (e.g., test-retest, parallel forms, internal consistency)
- **LO4** Evaluate reliability coefficients
- **LO5** Calculate and interpret standard error of measurement
- **LO6** Describe generalizability theory
- **LO7** Understand that instruments developed with item response theory calculate reliability differently

In Chapter 1, there was a discussion of the importance of examining an instrument's psychometric qualities to determine if it is appropriate for use. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), "in interpreting and using test scores, it is important to have some indication of reliability" (p. 33). **Reliability** is an area of assessment in which there have been some changes in the last 20 years, although how we have traditionally considered reliability (i.e., classical test theory) still plays an important role in how we examine the evidence related to reliability. The chapter title in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) related to reliability is "Reliability/Precision and Error of Measurement" (p. 33). As we will see, issues of *precision* and *error* are paramount in our consideration of reliability. According to the *Standards for Educational and Psychological Testing*, the term *reliability* has been used in two ways in the measurement literature. First, the term has been used to refer to reliability coefficients of classical test theory, which will be described in the next few pages. Second, reliability/precision has been referred to as the *consistency* of scores across replications of testing procedures that can be documented in various ways. Precision and consistency are important as,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

for example, a counselor would not want an instrument that indicated a client was moderately depressed once and then 10 minutes later indicated that the client was only minimally depressed. In general, there are three major theories related to reliability: classical test theory, generalizability theory, and item response theory. I will spend the majority of time discussing classical test theory, as historically it has had the most influence on how many instrument developers have conceptualized reliability.

L01 Classical Test Theory

Some people may have taken an assessment (e.g., a test for a class) and believed that the test did not adequately measure their knowledge in the area. For instance, a person may have felt physically ill or may have had a disagreement with someone just before the test. It might also be that the test questions were vague or not particularly well written. Some people, however, have performed better on a test than they originally thought they would. This could be because they guessed right on a couple of questions or perhaps the instructor did not ask many questions about the content covered the day they missed class. Whatever the case, the test scores did not reflect the individual's actual knowledge, or, to state it in another way, there was some *error* involved in the test scores. In terms of reliability, classical test theory is built on the premise that any result from an assessment device is a combination of an individual's true ability plus error (Crocker & Algina, 1986).

Although reliability is often considered an indicator of consistency, according to classical test theory, it is based on the degree to which there is error within the instrument. Classical test theory suggests that every score has two hypothetical components: a true score and an error component. Figure 3.1 presents this relationship with *Obs*. representing the

FIGURE 3.1
Classical test theory and reliability

- According to classical test theory, reliability is based on the concept that every observation (e.g., test score, instrument result, behavioral observation) is a combination of an individual's true score plus error:

$$Obs. = T + E$$

- Based on the above formula, theorists then proposed that the variance of the observation or score would be equal to the variance of the true score plus the variance of error:

$$s_O^2 = s_T^2 + s_E^2$$

- To make the above equation into something that can be used to determine reliability, we can divide the variance of observed on both sides of the equation:

$$\frac{s_O^2}{s_O^2} = \frac{s_T^2}{s_O^2} + \frac{s_E^2}{s_O^2}$$

- Because anything divided by itself is equal to 1, the equation then becomes

$$1 = \frac{s_T^2}{s_O^2} + \frac{s_E^2}{s_O^2}$$

- Theoretically, it is now possible to take the variance of error to the variance of observed off both sides of the equation:

$$1 - \frac{s_E^2}{s_O^2} = \frac{s_T^2}{s_O^2} + \frac{s_E^2}{s_O^2} - \frac{s_E^2}{s_O^2}$$

- Because anything minus itself is 0, the remaining equation is

$$1 - \frac{s_E^2}{s_O^2} = \frac{s_T^2}{s_O^2} = \text{reliability}$$

- Based on classical test theory, a measure of reliability provides an estimate of the amount of true variance to observed variance. Therefore, if an instrument's test manual indicates the instrument has a reliability coefficient of .80, this is interpreted, using classical test theory, as 80% of the variance is true to observed variance. In using the other side of the above equation, we can also see that 1 minus the error variance to observed variance equals reliability. Therefore, we can conclude that 20% of the variance is error variance (1 − .20 = .80).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

observed score; *T*, the true score; and *E*, the error. In the field of psychological and educational assessment, no one has been able to develop a perfect instrument; so all instruments are a combination of a true measure and some degree of error. The pertinent question is: How much of any score is true, and how much is error? Figure 3.1 shows how, in theory, we progress from the simple concept of an observed score being equal to a true score plus error to reliability being a ratio of true variance to observed variance. Therefore, when counselors read an instrument's manual that includes a reliability coefficient, they often use classical test theory to interpret that coefficient. For example, a reliability coefficient of .75 would indicate that 75% of the variance is true to observed variance. As Figure 3.1 also indicates, 1 minus the ratio of error variance to observed variance is equal to reliability ($1.00 - x = .75$). Therefore, we can determine that the ratio of error variance to observed variance is .25 ($1.00 - .25 = .75$). Thus, we use a reliability coefficient to give us an *estimate* of how much of the variance is true variance and how much is error variance.

Reliability and Random Error

In classical test theory terms, a reliability coefficient is an estimate of the proportion of total variance that is true variance and the proportion that is error variance. Error variance, in this case, is only *random error*; it does not include systematic error. Reliability is frequently calculated using the consistency of scores obtained by the same individuals, whereas variations in scores provide an estimate of error. Examining the difference between systematic and random error should clarify this distinction. *Systematic* means there is a system—methods are planned, orderly, and methodical. *Random error* means the lack of a system—occurrences are presumed to be unsystematic. An example of systematic error is a test question that contains a typographical error and everyone who takes the test reads that same error. Reliability would not measure this error because it is systematic. Systematic errors are constant errors that do not fluctuate.

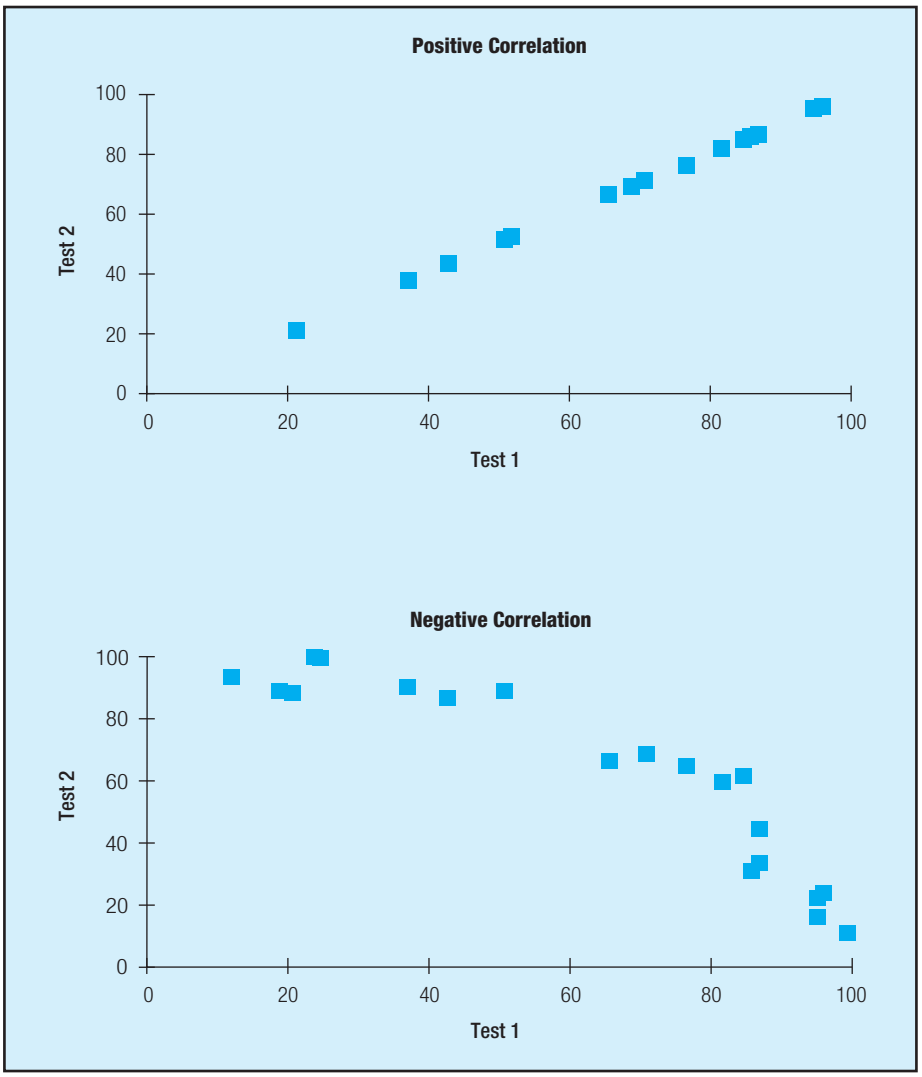
Random errors are those errors that are not consistent, such as a coffee spill on just one person's test. Random errors can arise from a variety of sources. Sometimes there are situations in which an error occurs in the administration of an instrument. For example, a school counselor responsible for the administration of the SAT at one school may not read the instructions correctly to students, whereas students at other schools receive the correct instructions. Illness, fatigue, and emotional factors can all affect an individual's performance on an assessment instrument and contribute to random error. Scores may also be affected by such factors as the lack of sleep, lucky guesses, or test anxiety. There may be problems with the facilities during the administration of an instrument (e.g., the room may be too hot, cold, or noisy). The examinee's performance might be affected because he or she is starting to get sick and has a terrible headache, which may result in random error. The construction of the items on the instrument may be vague, and people may vary in how they interpret the items. With any instrument, it is important to know the degree to which random error is a problem, and reliability coefficients provide an estimate of the proportion of random error. In the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), the authors use the term *reliability coefficients* to refer to the coefficients calculated using classical test theory, whereas they use the term *reliability/precision* to denote the more general notion of consistency across testing procedures.

LO2 Correlation Coefficients

With classical test theory, reliability is frequently *calculated* based on the degree of consistency between two sets of scores (e.g., give a group an instrument once, wait a period of time, and then test them again to see if the scores change). Often when individuals want to examine consistency, they use the statistical technique of correlation. Correlation provides

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 3.2
Scatterplots for positive and negative correlations



an indication of consistency by examining the relationship between scores. Figure 3.2 shows an example of a small class that took a test once and then took it again a week later. In this example, there is a perfect positive correlation (+1.00) because everyone in the class got the same score and the same rank the second time as they did in the first administration. The second scatter diagram in Figure 3.2 reveals a negative correlation—the scores on the *x*-axis increase, whereas the scores on the *y*-axis decrease. In some cases, we do not want a negative correlation. For example, with reliability coefficients, we would not want those who had the highest scores the first time they took the test to have the lowest scores the second time they took the test; nor would we want the reverse, where those who had the lowest scores at first have the highest scores the second time. There are occasions, however, when a negative correlation is expected. For example, as the number of counseling sessions increases, we would expect clients' levels of depression to decrease.

The statistical technique of correlation produces a **correlation coefficient**, which provides a numerical indicator of the *relationship* between the two sets of data. Remember that correlation is a statistic that is used not only to determine reliability, but it has multiple

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

uses that involve examining the relationship between two sets of variables. Correlation coefficients can range from −1.00 to 1.00. The closer the coefficient is to 1.00 (either positive or negative), the stronger the relationship. On the other hand, the closer the coefficient is to .00, the smaller the relationship. A correlation coefficient of .00 indicates the lack of evidence of a relationship. Correlation coefficients can be computed in various ways, and it is the nature of the data that determines the appropriate method.

The most common method for calculating correlations is the *Pearson-Product Moment Correlation Coefficient*, which is essentially based on the simple concepts of a positive times a positive is a positive, a negative times a negative is a positive, and a positive times a negative is a negative; a big times a big is a big, and a small times a small is a small. Figure 3.3 shows how these concepts are used to calculate a correlation coefficient. The formula for the Pearson-Product Moment Correlation is

$$r = \frac{\sum z_1 z_2}{N}$$

In this equation, r represents the Pearson-Product Moment Correlation and characterizes the sum of the product of the z scores (take each person's z score on the first administration, multiply it by the z score on the second test, and then add these products together). In this equation, N represents the number of individuals. Converting each individual's scores into z scores provides an indication of whether they are above or below the mean. Remember that z scores

FIGURE 3.3
Calculating a Pearson-Product Moment Correlation

	First Administration		Second Administration	
	Raw Scores	***z* Scores**	**Raw Scores**	***z* Scores**
Joseph	62	−1.90	80	−.78
Hannah	78	−.38	78	−1.10
Geraldo	75	−.67	91	.94
Megan	84	.19	85	.00
Seth	88	.57	87	.31
Maria	92	.95	78	−1.10
Jason	95	1.24	96	1.73
	$M = 82$		$M = 85$	
	$s = \sqrt{\frac{774}{7}} = \sqrt{110.57} = 10.52$		$s = \sqrt{\frac{284}{7}} = \sqrt{40.57} = 6.37$	

Product of *z* Scores
$-1.90 \times -.78 = 1.48$
$-.38 \times -1.10 = .42$
$-.67 \times .94 = -.63$
$.19 \times .00 = .00$
$.57 \times .31 = .18$
$.95 \times -1.10 = -1.05$
$1.24 \times 1.73 = 2.15$
2.55

$$r = \frac{\sum z_1 z_2}{N} = \frac{2.55}{7} = .36$$

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

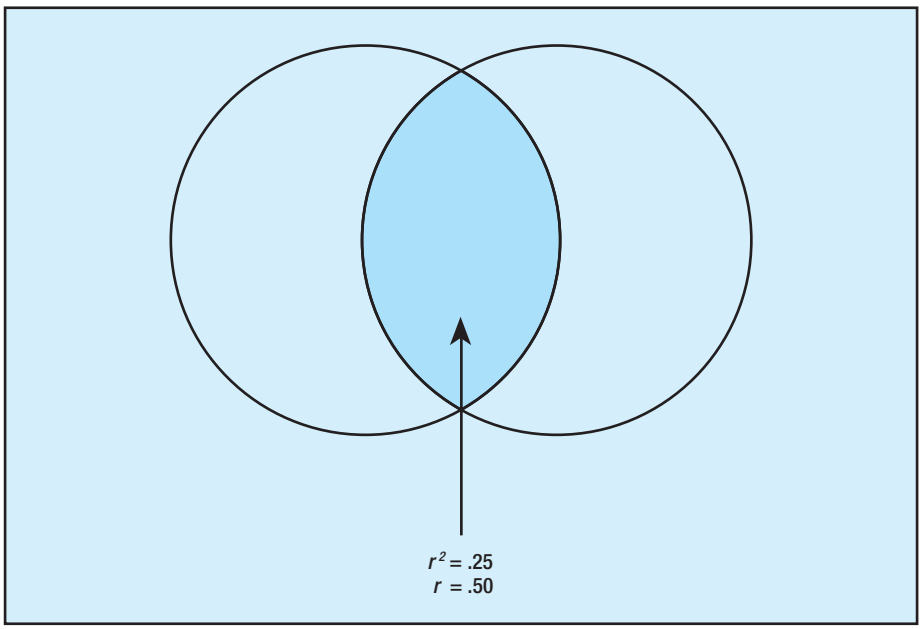
above the mean are positive numbers and *z* scores below the mean are negative numbers. We would expect each test taker to be consistent; in other words, each would be either above the mean or below the mean on both administrations (a positive times a positive and a negative times a negative is a positive). Thus, if test takers were consistent in their performance, that would increase the product of the *z* scores and, thus, increase the correlation coefficient. On the other hand, if they are above the mean one time (a positive *z* score) and below the mean the second time (a negative *z* score), then a positive times a negative is a negative, which would decrease the correlation coefficient. The same applies to the size of the *z* scores because a higher *z* score (a big) times a higher *z* score (a big) will produce a larger number than a higher *z* score times a smaller *z* score. What is important to understand is that consistent scores contribute to a larger correlation coefficient (toward either a −1.00 or a +1.00), whereas inconsistent scores reduce the size of the correlation coefficient (toward .00).

There are easier methods for calculating a Pearson-Product Moment Correlation that do not require converting every raw score into a *z* score. The following formula is easier to calculate and will be less time consuming:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$$

Even though this formula does not appear to take into consideration the consistency of raw scores, it actually does, because all the methods of calculating correlations take into account whether the two sets of data vary in the same manner. Sometimes with correlations, individuals want more than just an indicator of the relationship; they would like to know the percentage of common variances between the two sets of data. To find this percentage of shared variance between the two variables, we simply square the correlation coefficient. This statistic is called the **coefficient of determination** and is represented by r^2. Therefore, if a correlation coefficient is equal to .50, the amount of shared variance is equal to .25 ($r^2 = .50^2$). The amount of shared variance is often represented in a Venn diagram (see Figure 3.4), which visually depicts the amount of overlap between the two variables.

FIGURE 3.4
Venn diagram representing percentage of shared variance



Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Hence, if there were a large degree of shared variance, there would be a large section where the two circles overlap; however, if r^2 were quite small, then the circles would hardly intersect.

L03 Types of Reliability Coefficients

The previous discussion of correlation is necessary for understanding how an estimate of reliability is frequently calculated. As you will remember, in classical test theory, reliability coefficients provide us with an indication of the amount of true variance to observed variance. In everyday counseling, however, it is impossible to get a measure of "true" variance. Therefore, the statistical technique of correlation is often used to provide us with an estimate of reliability that is based on the concept of consistency (i.e., we would trust that a score is a "true" estimate if a person gets that same score consistently). The following discussion provides an overview of the typical methods for estimating reliability coefficients. Counselors need to understand the different types of reliability estimates as they relate to different sources of measurement error and, contrary to some perceptions, these estimates of reliability are not interchangeable (Dimitrov, 2007).

Test-Retest

A common method for calculating a reliability coefficient is to give the identical instrument twice to the same group of individuals. With the **test-retest reliability** method, a reliability coefficient is calculated simply by correlating the performance on the first administration with the performance on the second administration. It is expected that the variation in scores between the first and second administrations would relate to random error because everything else should be the same. The correlation coefficient between the test and retest scores provides us with an estimate of reliability. We can then use classical test theory to interpret the coefficient. Therefore, if the reliability coefficient using test-retest were .85, using classical test theory we would interpret it by saying that 85% of the variance is true to observed variance and 15% is error to observed variance. In reliability, we do not square the correlation coefficient or use the coefficient of determination. Instead, we simply use the reliability coefficient itself to evaluate the degree of measurement error.

Using test-retest as a method for documenting reliability does present some difficulties. One of the major problems is the possibility that something could intervene between the first administration and the second administration. For example, consider a counselor who is interested in developing an instrument to assess clients' potential for having an eating disorder. This counselor gave the test once, but before he could give it again, there was a popular movie dealing with eating disorders that was just released and many of the group members went to this movie. This movie would probably affect the clients' scores on the retest. Instrument manuals should specify the time period between the first and second administrations in a test-retest situation. In general, reliability coefficients tend to decrease as the amount of time between the testing increases. Anastasi (1988) recommended that the time between the test and retest should rarely exceed six months. Sometimes a test-retest reliability coefficient will be called a *coefficient of stability* when there is a long period between the administrations of the instrument. In this case, the test-retest reliability provides an indication of the stability of the scores.

The test-retest method of estimating the reliability of an instrument is appropriate only if the following assumptions are met. First, the characteristic or trait that is being measured must be stable over time. If an instrument measured situational anxiety (not trait anxiety), we would anticipate much more fluctuation in scores, and the test-retest method would not be appropriate. Trait anxiety, however, is supposedly stable over time, so, in this case,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

test-retest would be an appropriate method for estimating reliability. The second assumption is that there should be no differential in practice effect. With many skills, people become more proficient as they practice, and the first administration of the test could serve as that practice. For example, in a psychomotor test where individuals are required to screw nuts and bolts together, individuals may improve with practice and score significantly higher on the retest. The third, and last, assumption is that there should be no differential in learning between the test and the retest. Exposure to education or psychological interventions between the testings can influence the performance on the second test. Sometimes learning can occur as a result of the first test. For example, sometimes people will simply remember the correct response to an item if the time between testings is fairly short. In addition, it is not uncommon for people to be bothered by a question and, after the test is over, seek the correct answer to that question. Some people may thus learn as a result of taking the first test. Therefore, there are instruments for which test-retest is appropriate to use and others for which test-retest is not appropriate. As consumers of assessment instruments, counselors need to decipher when the test-retest method is an appropriate technique for determining a reliability coefficient for an instrument.

Alternate or Parallel Forms

Calculating a reliability coefficient using **alternate or parallel forms** requires that there be *two* forms of an instrument. In this method, individuals are given one form of the instrument initially and then are assessed with a second alternate or parallel form of the instrument. The scores on the two instruments are then correlated, which results in an estimate of reliability. This method avoids the many difficulties of the test-retest method because it uses different forms of the instrument. Because the two forms eliminate the problems associated with individuals remembering responses, the forms can be given in immediate succession.

Care should be taken with alternate forms to ensure that the two forms are truly parallel. Ideally, the two forms of the instrument should be independently constructed to meet the same test specifications. For the instruments to be equivalent, there must be the same number of items, and the items must be in the same format. Content sampling becomes critical because both forms must adequately sample the content domain equally. Most content domains will vary in terms of difficulty, so instrument developers need to guard against one form being more difficult than the other. In the personality area, it is important to ensure that similar questions are addressing the same issue; even subtle wording changes can alter the context in which individuals answer questions. For example, some people would read the questions "Do you fight with your spouse?" and "Do you argue with your spouse?" as being equivalent. Others, however, may interpret *fight* as being physical and *argue* as being verbal. In the instrument's manual, there should be substantial evidence that instrument developers made substantial efforts to ensure the forms are truly equivalent.

Although the alternate-form method of calculating a reliability coefficient offers many advantages, there are also some difficulties with using this method. Often in the area of counseling, it is difficult to develop even one sound instrument; the efforts required to develop two different but equivalent forms can be monumental. As a result, the alternate or parallel forms are a less prevalent method of estimating reliability. In addition, with alternate forms, the first instrument may have some influence on the administration of the second instrument (e.g., practice effect or learning) that will affect the reliability coefficient.

Internal Consistency Measures of Reliability

The following methods for calculating reliability coefficients all use one administration and a single form of the instrument. Each method divides the instrument in different manners and correlates the scores from the different portions of the instrument. Thus, these forms of reliability examine the instrument "internally" to determine its consistency or lack of error.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Split-Half Reliability. In the **split-half reliability** method, the instrument is given once and then split in half to determine the reliability. The first step in this method involves dividing the instrument into equivalent halves. Splitting the instrument into the first half and the second half is sometimes not appropriate because some tests become progressively more difficult. In addition, warming up and practice effects may influence performance on the first half, whereas fatigue and boredom may influence performance on the second half. Frequently, an instrument is divided in half by using the scores from the even items and the scores from the odd items. This odd–even splitting often minimizes some of the problems with equivalency, particularly if the items are arranged by order of difficulty. Other methods for dividing an instrument involve selecting items randomly or organizing the items on the instrument by content.

After the instrument has been divided into two halves, the second step involves correlating each individual's scores on the two halves. This is problematic, however, because a correlation coefficient is influenced by the number of observations involved in the calculation (i.e., in general, a larger number of observations will produce larger correlation coefficients and a smaller number of observations will produce smaller correlation coefficients). Therefore, when we split an instrument in half and correlate the halves, we are using half the original number of items in the calculation. For example, if an instrument involves 50 items, the split-half method provides an estimate of reliability based on 25 items. Because longer instruments usually have higher reliability coefficients than shorter instruments, split-half reliability coefficients will be smaller than those of other methods that use the original number of items. Research has shown that the **Spearman–Brown formula** provides a good estimate of what the reliability coefficient would be if the halves were increased to the original length of the instrument. Instrument developers often use the Spearman–Brown formula to correct split-half reliability estimates. The Spearman–Brown formula is

$$r_{ii} = \frac{2r_{hh}}{1 + r_{hh}}$$

where r_{ii} represents the Spearman–Brown correction of the split-half reliability coefficient and r_hh is the original split-half correlation coefficient. When comparing the reliabilities of different instruments, a split-half reliability coefficient should never be directly compared with a Spearman–Brown reliability coefficient, because it would be expected that the Spearman–Brown would always be higher. As an example, would you select the ABC Self-Esteem Inventory with a split-half reliability coefficient of .88 or the XYZ Self-Esteem Inventory with a Spearman–Brown reliability coefficient of .88? Because the Spearman–Brown is a correction of the split half, you should not assume that these two instruments are equally reliable. Therefore, you would be correct in selecting the ABC Self-Esteem Inventory because the split-half reliability coefficient of .88 would be increased if the Spearman–Brown formula were used.

Kuder–Richardson Formulas (KR-20 and KR-21). The two **Kuder–Richardson formulas** also enable you to calculate reliability by examining the internal consistency of one administration of an instrument. The decision to use either Kuder–Richardson formula 20 (KR-20) or Kuder–Richardson formula 21 (KR-21) is based on the characteristics of the instrument, specifically whether the items are measuring a homogeneous or heterogeneous behavior domain. To see the difference between homogeneous and heterogeneous instruments, let's use the examples of instruments that measure family functioning. An example of a homogeneous instrument would be an instrument that measures family functioning and assesses only the single domain of family communication. On the other hand, with a heterogeneous instrument, multiple domains are examined, such as an instrument that measures family communication, family cohesion, family adaptability, and family conflict. The KR-20, which is purported to be an estimate of the average of all split-half reliabilities

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

computed from all the possible ways of splitting the instruments into halves, is appropriate for instruments that are heterogeneous. The formula for computing a KR-20 is

$$r_{20} = \left(\frac{k}{k-1}\right)\frac{s_t^2 \Sigma pq}{s_t^2}$$

where r_{20} is the reliability coefficient of the instrument using the KR-20 formula, k is the number of items on the instrument, and s_t^2 is the standard deviation of the instrument squared, which is the variance of the instrument. The terms p and q may be new to you. We sum (Σ) the proportion of individuals getting *each* item correct (p) multiplied by the proportion of individuals getting each item incorrect (q). With the KR-20, one must calculate for each item the number of people who got it right and the number of people who got it wrong. Computers can be of assistance in calculating KR-20s because this method can be quite time consuming.

The formula for the KR-21 is much simpler to compute, but it is appropriate only for an instrument that is *homogeneous.* The formula for the KR-21 is

$$r_{21} = \left(\frac{k}{k-1}\right)\left[1 - \frac{M(k-M)}{ks^2}\right]$$

where r_{21} is the reliability coefficient using the KR-21 formula, k is the number of items, M is the mean, and s is the standard deviation. Sometimes you may see an instrument manual in which the instrument developers report a KR-21 because it is easy to calculate when the developers should have used a KR-20. The KR-21 cannot be used if the items are not from the same domain or if the items differ in terms of difficulty level. In general, the Kuder–Richardson formulas tend to have lower reliability coefficients than the split-half method.

Coefficient Alpha. In terms of internal consistency methods of estimating reliability, the discussion has so far concerned only examples where the scoring is dichotomous (e.g., right or wrong, true or false, like me or not like me). Some instruments, however, use scales that have a range, such as strongly agree, agree, neutral, disagree, and strongly disagree. On some personality tests, the answers to questions have different weightings. When the scoring is not dichotomous, then the appropriate method of estimating reliability is **coefficient alpha** or **Cronbach's Alpha**. Coefficient alpha (α) takes into consideration the variance of each item. (If you would like to calculate a Cronbach's Alpha, see Cronbach, 1951.) Coefficient alphas are usually low and conservative estimates of reliability.

Nontypical Situations

Typical methods of calculating reliability coefficients are not appropriate for some types of assessment tools. As consumers of assessment instruments, counselors need to have an understanding of the assessment areas in which special consideration should be taken to determine the reliability of an instrument. One area in which typical methods of determining reliability are not appropriate is with speed tests, such as a psychomotor test where examinees are assessed to see how many nuts and bolts they can screw together in a specified time period. A test-retest also may not be appropriate, because many individuals will become better at this task with practice. Also, splitting these timed tests in half or using other measures of internal consistency would probably result in nearly perfect reliability. For example, if an examinee got 22 nuts and bolts together in the allotted time, she would get 11 even "items" correct and 11 odd "items" correct, which would result in an overestimate of the instrument's reliability. Sometimes, to address this problem, instruments are split in half and people are given a certain time period to complete each half; in this case, the number of items in each timed period is correlated.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Another type of instruments in which typical methods for determining reliability may be problematic is with criterion-referenced instruments that involve mastery. As you will remember, criterion-referenced instruments compare an individual's performance with a standard or criterion. For example, a teacher may want to determine whether students have mastered multiplying one-digit numbers before allowing them to proceed with learning how to multiply two-digit numbers. Another criterion-referenced instrument could be one in which there is a cutoff score and all scores above that point would indicate that a client is considered to be depressed. With these criterion-referenced instruments, we would expect that most of the scores would be close to the previously determined criterion or standard. In fact, in some situations, all of the examinees continue to receive instruction until the mastery score is achieved. For example, a teacher might require that all students show mastery of one-digit multiplication by getting 90% of the items on a test correct. Therefore, there is often little variability in criterion-referenced instruments. Remember the discussion of correlation and the phrase "a big times a big is a big, a small times a small is a small"? This phrase applies to criterion-referenced instruments because there will be only small variations from the criterion score; therefore, the correlation coefficients will be small due to the lack of variability. Typical methods of determining reliability use correlation, but these methods will frequently result in low reliability coefficients for criterion-referenced instruments. Dimitrov (2007) contended that reliability evidence for criterion-referenced instruments is often provided by an indicator of *classification consistency.* In this case, the instrument may be given twice and a statistical analysis of consistency is conducted. If you are interested in examining the different methods of determining reliability for criterion-referenced instruments, the following resources may be helpful to you: Berk (1984) and Subkoviak (1984). In addition, it is difficult to determine estimates of reliability for those instruments that use item branching, where one's answer to one question will influence whether the next question is more difficult or, possibly, easier. Many of these types of tests use item response theory, which is briefly discussed later in this chapter.

Instruments can also vary in the degree to which the scoring is more objective or subjective. As later chapters discuss, there are instruments in which the scoring involves clinical judgment, which requires some subjectivity in scoring. With subjectively scored instruments, we are not only interested in whether examinees score consistently but also interested in examining any scoring differences among the examiners who score the same instrument. If the scoring of an instrument requires some professional judgments, then the instrument developers need to provide information on **interrater reliability**, which refers to how consistently different raters evaluate the answers to the items on the instrument. Although there are different methods for determining interrater reliability, the range is the same as for other reliability indices (.00 to 1.00). Interrater reliability becomes particularly important with portfolio assessments, behavioral observations, simulation exercises, and performance assessments, which will be discussed in later chapters.

L04 Evaluating Reliability Coefficients

The evaluation of an instrument's reliability/precision evidence should not be done in isolation. In instrument manuals, you will often find a range of reliability coefficients rather than just one coefficient. A counselor cannot just look at reliability coefficient(s) to determine whether there is adequate reliability. For instance, if you were to examine the reliability coefficient of .91, you might think that it was certainly acceptable. However, as will be discussed in a later overview of generalizability theory, reliability coefficients can convey quite different information. For example, a reliability coefficient could reflect various sources of error or possibly only one or two minimal sources of error. The manuals for most well-established instruments will include various reliability estimates for the user to evaluate. In evaluating reliability

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

coefficients, a counselor should not assume that there are preferred approaches to estimating reliability. As the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) indicates, one should not infer that test-retest or alternate-form reliability estimates are always preferable to estimates of internal consistency because the appropriate procedures for estimating reliability will depend on the type of assessment and its purpose.

In order to determine whether an instrument has adequate reliability to use, we must first examine the purpose for using the instrument. If the results of an instrument could have a substantial impact on an individual's life, then we would want an instrument with very high reliability coefficients. In many school districts, for example, intelligence or general ability tests are part of the information used to determine whether a child can receive special educational services. It would be unfortunate if children did not receive the services they needed because of an unreliable test. As you will see in Chapter 9, some of the widely used intelligence instruments have overall reliability coefficients that are slightly higher than .91. However, many personality instruments do not have reliability coefficients in the .91 range. Therefore, in evaluating reliability coefficients, counselors must examine how they are going to use the instrument and be knowledgeable about the reliability coefficients of other instruments in that area.

In selecting an instrument, counselors also need to examine the characteristics of their clients and analyze whether the reliability information is applicable to that population. Sometimes an instrument will have acceptable overall reliability but will not be particularly reliable with certain subgroups (e.g., African American clients). As a general rule, instruments that assess younger children are less reliable than instruments for adults. Hence, a counselor could make a mistake by using an instrument that is not particularly reliable for the age group with whom he is working. The reliability coefficients for different ethnic groups or socioeconomic groups may also vary with instruments. Therefore, a counselor cannot rely on, for example, a brochure promoting the instrument; instead, she must examine the manual and evaluate all of the information concerning the reliability/precision of the instrument.

LO5 Standard Error of Measurement

As previously stated, helping professionals should evaluate the reliability coefficients of different assessment tools when selecting appropriate assessment interventions. Reliability coefficients, however, is useful to practitioners not only in selecting and evaluating instruments but also in interpreting clients' scores. Using classical test theory, reliability is theoretically based on the concept of true score plus error. If a client took a test many times (e.g., 100 times), we would expect that one of those test scores would be the client's "true score." Hence, reliability is not something an instrument has but, rather, it is related to results and the degree of error in those results. **Standard error of measurement** provides an estimation of the range of scores that would be obtained if someone were to take an instrument over and over again. Although standard error of measurement is applicable to other theories of reliability than classical test theory, with classical test theory the focus is on identifying the range of scores where we would expect a client's "true score" to fall.

Let's examine a case of how standard error of measurement can be useful. I have often been consulted by parents about whether their child should retake a college entrance examination (e.g., SAT or ACT) to try to get better scores. Let's talk about the example of Patrick, whose scores on the SAT Critical Reading was 552. Patrick was planning on mainly applying to colleges that accepted students with scores typically above 600, and he wanted to know if he took the SAT again what were the chances his score would be above 600. Standard error of measurement can provide individuals information by informing them about where one could expect their "true score" if they were to take the examination

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

100 times. Based on Patrick's first result, we can expect that 95% of the time his Critical Reading score would fall between 492 and 612. Hence, Patrick may get a higher score of above 600 on the SAT Critical Reading test the next time he took it, but he also might get a lower score than the 552. Interestingly, Patrick retook the SAT and his Critical Reading score did slightly improve to a score of 565.

As indicated earlier, standard error of measurement can provide a range of where people would expect their scores to fall if they took the instrument repeatedly. Standard error of measurement is based on the premise that when individuals take a test multiple times, the scores will fall into a normal distribution. Thus, when individuals take an instrument numerous times, they tend to score in a similar fashion to their original score, whereas dissimilar scores are far less prevalent. In Figure 3.5, the three bell-shaped curves indicate the expected range of scores for George, Michael, and Hillary if they took an instrument multiple times. Figure 3.6 is an enlargement of Figure 3.5, where we can examine just Michael's expected range of scores. If Michael were to take an instrument multiple times, we would expect that 68% of the time his true scores would fall between 1 SEM below his original score and 1 SEM above his original score. In addition, we could expect that 95% of the time

FIGURE 3.5
George's, Michael's, and Hillary's range of scores using standard error of measurement

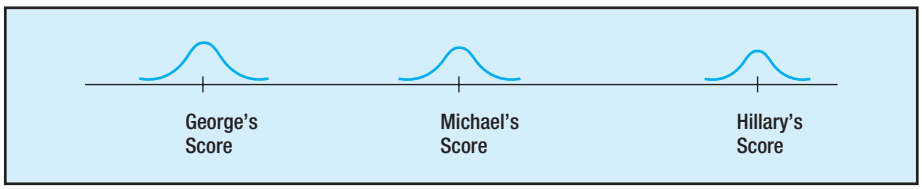
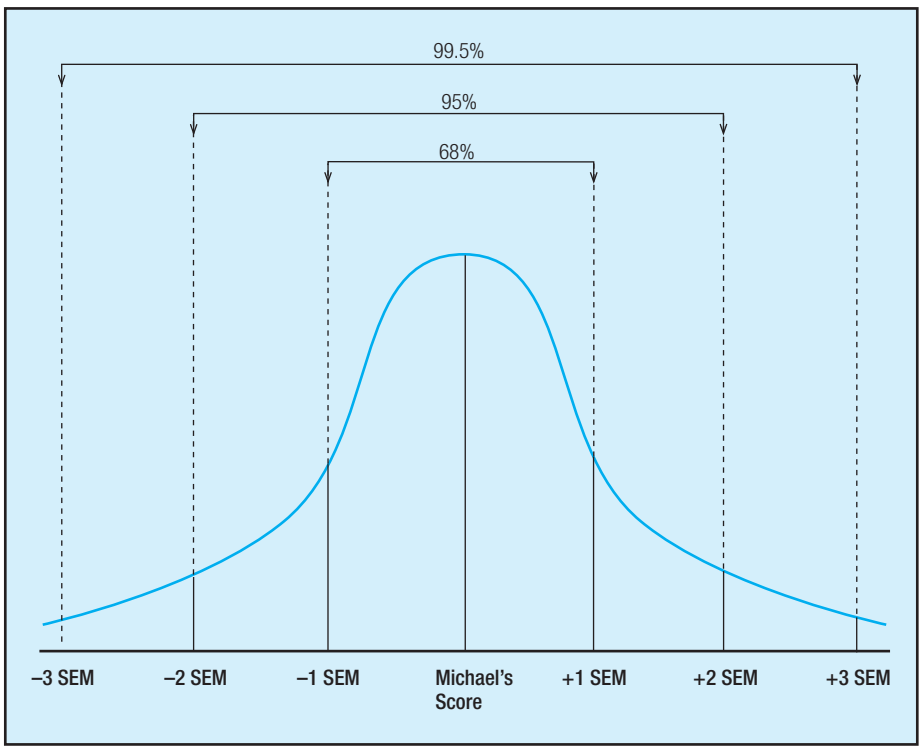


FIGURE 3.6
Range of Michael's scores using standard error of measurement



Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

his true score would fall between 2 SEM below and 2 SEM above his score, and 99.5% of the time his true score would fall between 3 SEM below and 3 SEM above his score.

The formula for calculating the standard error of measurement is:

$$SEM = s\sqrt{1 - r}$$

where *s* represents the standard deviation of the instrument and *r* is the *reliability coefficient*. Let's return to our example of Patrick and his score of 552 on the Critical Reading portion of the SAT. The scores on the Critical Reading portion of the SAT range from 200 to 800. Often, the standard error of measurement is provided in the instrument's manual, so the counselor does not need to calculate it. In this case, however, we will calculate the standard error of measurement as an example. Although means and standard deviations for the SAT scores vary, we will say that for this example the mean is 500 and the standard deviation is 100. The reliability coefficients for Critical Reading range from .91 to .93 (College Board, 2010). For this example, we will use the lowest reliability estimate (i.e., .91). Therefore, the standard error of measurement would be $100\sqrt{1.00 - .91} = 100\sqrt{.09} = 100(.30) = 30$. We would then add and subtract the standard error of measurement to Patrick's score to get the range. A counselor could then tell Patrick that 68% of the time he could expect his Critical Reading score to fall between 522 (552 − 30) and 582 (552 + 30). If we wanted to expand this interpretation (which we did earlier), we could use two standard errors of measurement ($2 \times 30 = 60$). In this case, we would say that 95% of the time Patrick's score would fall between 492 (552 − 60) and 612 (552 + 60). If we wanted to further increase the probability of including his true score, we would use three standard errors of measurement ($3 \times 30 = 90$) and conclude that 99.5% of the time his true score would fall between 462 (552 − 90) and 642 (552 + 90). Therefore as you see, standard error of measurement can be used to create confidence intervals around the client's score (AERA, APA, & NCME, 2014).

Standard error of measurement incorporates reliability coefficients in determining the range of scores. To illustrate this point, let's say there are two tests that measure intelligence, and both instruments have the same mean (100) and standard deviation (15). Juan took these two instruments and obtained the same IQ score of 104 on both instruments. Because Juan obtained the same score, we might expect that the range of where we would estimate Juan's true score to fall would be similar, but this is not the case because the reliability coefficients for the two instruments vary. In the first instrument, the reliability coefficient for the overall IQ measure is .90, whereas the reliability coefficient for the second IQ test is .71. Therefore, if we calculate the standard error of measurement for the first instrument, we get $15\sqrt{1.00 - .90} = 4.74$. With this first intelligence test, Juan could expect that 68% of the time his IQ score would fall between 99.26 and 108.74. On the second intelligence test, the SEM would be $15\sqrt{1.00 - 0.71} = 8.08$. On this test, Juan could expect that 68% of the time his IQ score would be between 95.92 and 112.08. As we see, the range, or score band, is notably larger for the second test because the instrument is less reliable. Standard error of measurement will always increase as reliability decreases.

An instrument's reliability can be expressed in terms of both the standard error of measurement and the reliability coefficient. The standard error of measurement is more appropriate for interpreting individual scores, and the reliability coefficient can be used to compare different instruments. Providing clients with a possible range of scores is a more proficient method of disseminating results than reporting only one score. Reporting a single score can be misleading because no instrument has perfect reliability. In the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

cognitive realm, a range of scores is a better indicator of a client's capabilities. People also vary in their personality and other affective dimensions, so using standard error of measurement can assist clients in understanding that variation. For example, a client may not be depressed on the day the results are interpreted and may dismiss information if the counselor provides only a single score indicating moderate depression. Presenting the results as a range takes into consideration the fluctuations in depression and increases the probability that the client will find the results useful. Counselors should be encouraged, whenever possible, to interpret assessment results by using a range of scores based on standard error of measurement. Because many instrument publishers are aware of the importance of reporting results using standard error of measurement, they are beginning to report results using score bands rather than single scores.

Although test scores have historically been used to evaluate and provide information about individuals, recently the use of group scores has increased (e.g., achievement scores for a school) to evaluate a program or a school's effectiveness. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) reflects that the error factors operating within the interpretation of group scores is different from those error factors influencing individual scores. The *Standards,* therefore, indicates that standard error of measurement should not be used for interpretation of group scores; instead, the more appropriate statistic is the standard error of the group means.

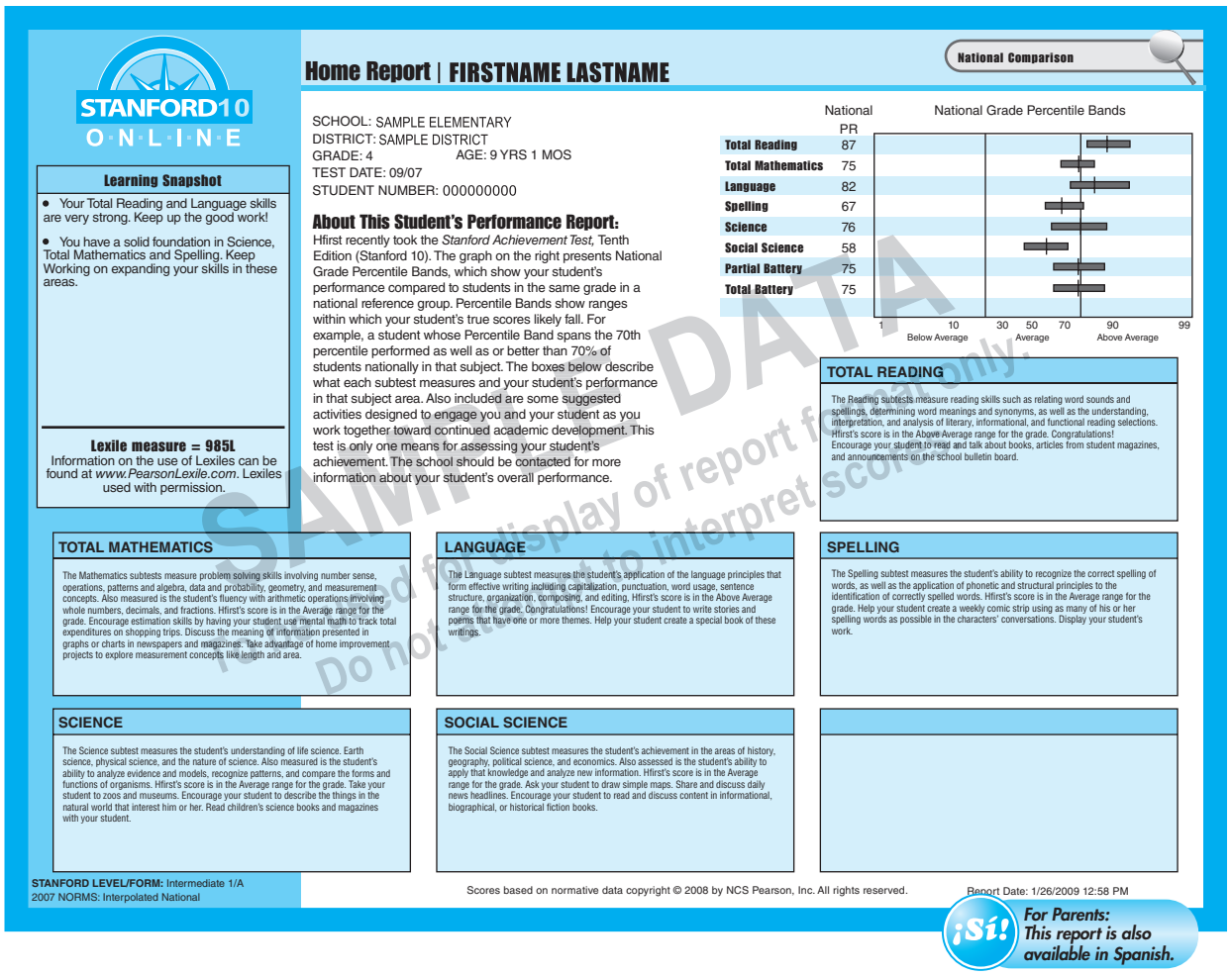
Standard Error of Difference

The purpose of some assessment situations is to compare certain aspects within the client. Often in counseling, for instance, a counselor wants to identify a client's strengths and limitations. Sometimes the focus concerns whether the individual has more aptitude in the verbal area as compared with the numerical area. On an achievement test, parents may want to know if their child's score in reading is significantly lower than in other academic areas. Examining differences in factors within a client also applies to personality assessment. Figure 3.7 illustrates percentile bands provided in the scoring profile of the Stanford Achievement Test. In this profile, the band around the percentile score represents a range from 1 SEM below the percentile score to 1 SEM above. With this profile, the counselor can examine strengths and limitations in terms of academic achievement. In examining Figure 3.7, you will notice that some of the percentile bands overlap, whereas there is no overlap among other percentile bands. If the bands overlap, you should *not* conclude that these scores are different. Differences can be attributed to areas where there is no overlap between the bands. With some instruments, the standard errors of measurement are not visually presented, so a counselor will need to compute the point difference to determine if scores are truly different. In examining differences between two scores, counselors need to take into account the errors of measurement from *both* scores. Examining the standard error of measurement from just one score will not be sufficient to determine the amount of difference actually needed to indicate a meaningful distinction. The **standard error of difference** enables counselors to calculate the exact amount of difference needed between scores. Standard error of difference takes into consideration the standard errors of measurement of both of the scores, as indicated in the following formula:

$$SE_{diff} = \sqrt{SEM_1^2 + SEM_2^2}$$

In this formula, SE_{diff} is the standard error of differences, SEM_1^2 is the standard error of measurement for the first score squared, and SEM_2^2 is the standard error of measurement for the second score squared.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.



STANFORD10 ONLINE

Home Report | FIRSTNAME LASTNAME

National Comparison

SCHOOL: SAMPLE ELEMENTARY
DISTRICT: SAMPLE DISTRICT
GRADE: 4 AGE: 9 YRS 1 MOS
TEST DATE: 09/07
STUDENT NUMBER: 000000000

Learning Snapshot

- Your Total Reading and Language skills are very strong. Keep up the good work!
- You have a solid foundation in Science, Total Mathematics and Spelling. Keep Working on expanding your skills in these areas.

Lexile measure = 985L
Information on the use of Lexiles can be found at *www.PearsonLexile.com*. Lexiles used with permission.

About This Student's Performance Report:
Hfirst recently took the *Stanford Achievement Test,* Tenth Edition (Stanford 10). The graph on the right presents National Grade Percentile Bands, which show your student's performance compared to students in the same grade in a national reference group. Percentile Bands show ranges within which your student's true scores likely fall. For example, a student whose Percentile Band spans the 70th percentile performed as well as or better than 70% of students nationally in that subject. The boxes below describe what each subtest measures and your student's performance in that subject area. Also included are some suggested activities designed to engage you and your student as you work together toward continued academic development. This test is only one means for assessing your student's achievement. The school should be contacted for more information about your student's overall performance.

	National PR
Total Reading	87
Total Mathematics	75
Language	82
Spelling	67
Science	76
Social Science	58
Partial Battery	75
Total Battery	75

National Grade Percentile Bands
1 10 30 50 70 90 99
Below Average Average Above Average

TOTAL READING
The Reading subtests measure reading skills such as relating word sounds and spellings, determining word meanings and synonyms, as well as the understanding, interpretation, and analysis of literary, informational, and functional reading selections. Hfirst's score is in the Above Average range for the grade. Congratulations! Encourage your student to read and talk about books, articles from student magazines, and announcements on the school bulletin board.

TOTAL MATHEMATICS
The Mathematics subtests measure problem solving skills involving number sense, operations, patterns and algebra, data and probability, geometry, and measurement concepts. Also measured is the student's fluency with arithmetic operations involving whole numbers, decimals, and fractions. Hfirst's score is in the Average range for the grade. Encourage estimation skills by having your student use mental math to track total expenditures on shopping trips. Discuss the meaning of information presented in graphs or charts in newspapers and magazines. Take advantage of home improvement projects to explore measurement concepts like length and area.

LANGUAGE
The Language subtest measures the student's application of the language principles that form effective writing including capitalization, punctuation, word usage, sentence structure, organization, composing, and editing. Hfirst's score is in the Above Average range for the grade. Congratulations! Encourage your student to write stories and poems that have one or more themes. Help your student create a special book of these writings.

SPELLING
The Spelling subtest measures the student's ability to recognize the correct spelling of words, as well as the application of phonetic and structural principles to the identification of correctly spelled words. Hfirst's score is in the Average range for the grade. Help your student create a weekly comic strip using as many of his or her spelling words as possible in the characters' conversations. Display your student's work.

SCIENCE
The Science subtest measures the student's understanding of life science. Earth science, physical science, and the nature of science. Also measured is the student's ability to analyze evidence and models, recognize patterns, and compare the forms and functions of organisms. Hfirst's score is in the Average range for the grade. Take your student to zoos and museums. Encourage your student to describe the things in the natural world that interest him or her. Read children's science books and magazines with your student.

SOCIAL SCIENCE
The Social Science subtest measures the student's achievement in the areas of history, geography, political science, and economics. Also assessed is the student's ability to apply that knowledge and analyze new information. Hfirst's score is in the Average range for the grade. Ask your student to draw simple maps. Share and discuss daily news headlines. Encourage your student to read and discuss content in informational, biographical, or historical fiction books.

STANFORD LEVEL/FORM: Intermediate 1/A
2007 NORMS: Interpolated National

Scores based on normative data copyright © 2008 by NCS Pearson, Inc. All rights reserved.

Report Date: 1/26/2009 12:58 PM

¡Sí! For Parents: This report is also available in Spanish.

SAMPLE DATA
Intended for display of report format only.
Do not attempt to interpret scores.

FIGURE 3.7
Illustration of score bands from the Stanford Achievement Test. Sample Home Report from the *Stanford Achievement Test Series, Tenth Edition* (Stanford 10). Copyright © 2003 NCS Pearson, Inc. Reproduced with permission. All rights reserved.

LO6 Generalizability Theory

Thus far, the examination of reliability/precision has only addressed the classical test theory model and the use of reliability coefficients to interpret clients' scores. Although this model continues to be influential, a discussion of reliability/precision should also include an exploration of alternative models. Whereas classical test theory considers a single distribution of errors in a test taker's scores, generalizability theory attempts to evaluate the contribution of various sources of error (AERA, APA, & NCME, 2014). In **generalizability theory** or **domain sampling theory**, the counselor is interested in the dependability, or the "generalizability," of the scores. With a true score approach, the focus is on identifying the portion of the score that is due to true score variance as compared with error variance. In generalizability or domain sampling, however, the focus is on estimating the extent to which specific sources of variation under defined conditions contribute to the score on the instrument. According to Cronbach (Cronbach, 1990; Cronbach, Gleser, Rajaratnam, & Nanda, 1972), the same instrument given under the exact same conditions should, theoretically, result in the exact same score. With generalizability theory, the focus is on the degree to which scores can be generalized across alternative forms, test administrators, number of items, and other facets of assessment. Therefore, when reliability information is

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

gathered, there needs to be a description of the "universe." Instead of using the term *true score,* the term in generalizability theory is *universe score*, which represents the average of taking the instrument multiple times in the same universe (e.g., same purpose and same conditions). When counselors use a single score and assume it represents a universe of scores, they are generalizing from that score. Cronbach preferred the word *generalize* to *reliability* because the former implies the act of generalizing to a specific purpose rather than an all-inclusive evaluation of the instrument.

Thus, generalizability theory requires multiple reliability coefficients or *coefficients of generalizability.* This theory encourages the study of variations in test scores with the intent of identifying sources of variation. For example, researchers would attempt to identify the portions of the variance that are due to practice effect, examinees' levels of motivation, administrators' styles, and room temperature. Generalizability theory directs us to consider that reliability is not a stagnant notion; instead, reliability varies depending on the purpose, setting, conditions, and use of the instrument. The interest is still in determining the degree to which the instrument is free from error. In fact, using this theoretical approach, a researcher would attempt to identify sources of error and, when it is appropriate, to generalize from the assessment results.

Vacha-Haase (1998) developed a meta-analytic method to explore variability in reliability estimates called **reliability generalization**. Similar to other meta-analytic procedures, reliability generalization involves combining reliability estimates across studies, which allows researchers to characterize and explore variance in score reliability. Vacha-Haase's technique has spurred a number of reliability generalization studies of instruments. A few examples of reliability generalization are: for the Dyadic Adjustment Scale (Graham, Liu, & Jeziorski, 2006), the Myers–Briggs Type Indicator (Capraro & Capraro, 2002), the Minnesota Multiphasic Personality Inventory (Vacha-Haase, Tani, Kogan, Woodall, & Thompson, 2001), and the Maslach Burnout Inventory (Wheeler, Vassar, Worley, & Barnes, 2011).

Although counselors are encouraged to use standard error of measurement in interpreting client's scores, it is sometimes difficult to determine which reliability coefficient to use in calculating the standard error of measurement as often there are multiple reliability coefficients, which are encouraged by generalizability theory. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) suggests using a comprehensive reliability or generalizability coefficient that might be available from studies using reliability generalization. It is often difficult, however, to find a coefficient that incorporates all sources of error and it is suggested that clinicians use the most appropriate reliability or generalizability coefficient for that client. Often reliability or generalizability coefficients will vary depending on the level of the score. That is, scores that are very low or very high tend to be less reliable than scores that are more typical. **Conditional standard errors of measurement** take into account the different score levels. For example, the GRE uses conditional standard errors of measurement for the different scores on that instrument. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) suggests that conditional standard errors of measurement are often better than a single average standard error for a population that assumes that the standard error is constant across score levels.

LO7 Item Response Theory

Another alternative to classical test theory is Item Response Theory (IRT), which is discussed in Chapter 4. Let me start by saying IRT is not one approach to instrument development, but rather, it is a family of theories and methods that are being increasingly used to develop instruments. In general, IRT addresses the precision with which an observable task or performance on items can be used to estimate the value of a latent trait

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

for each test taker. So often in counseling, the constructs being measured are unobservable (i.e., latent); therefore, this group of theories is also called **latent-trait theory**. Here the focus is on each item; hence, traditional forms of calculating reliability/precision are not appropriate. Although IRT can be used in personality assessment, it is more frequently used in developing items to measure achievement levels and requires both psychometric and technical sophistication as various mathematical models can be used. IRT involves examining the characteristic curve of each item; thus, the evaluation of precision or error within the instrument is also quite complex and statistically sophisticated. Essentially in IRT, these item characteristic curves plot the probability estimates of the amount of ability or trait level needed to respond to each item in a certain way. According to Urbina (2014), IRT can be used in two reciprocal ways: (a) to estimate the probability that individuals with specific levels of an ability or a trait will answer a question correctly or respond in certain ways, and (b) to estimate the level of ability or trait level needed in order to have a specified probability of responding in a certain way to an item.

Although a thorough description of IRT is beyond the scope of this book, more information on it is contained in Chapter 4. Concerning reliability/precision, the focus on IRT is on item characteristic curves. If the appropriate data is available, then the parameters of each characteristic curve can be generated for each item on the test. The *test information function* can then be calculated from the parameter estimates. Although information functions are somewhat analogous to reliability coefficients and standard error of measurement, they are derived from a different theoretical stance and therefore should not be confused. However, because IRT often involves people from different abilities levels, it is often possible to calculate conditional standard error of measurement, which the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) indicates is often preferable to use rather than one single standard error of measurement for the entire instrument.

In conclusion, the field of measurement is continuing to refine methods and approaches to reliability/precision. These advances are critical to counselors because these researchers are advancing methods for assisting counselors in considering whether client scores are reliable. Although classical test theory is still influential, counselors should keep abreast of changes in the field of measurement and refinements related to reliability.

Summary

Reliability concerns the degree to which scores on an instrument are consistent and the degree to which those scores may be influenced by error. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) uses the term *reliability coefficients* to describe the coefficients produced from classical test theory and the term *reliability/precision* to denote the more general notion of consistency of the scores across testing procedures. Using classical test theory, we interpret a reliability coefficient as the ratio of true variance to observed variance. A reliability coefficient of .87 indicates that 87% of the variance is true variance. It is impossible, however, to measure true variance; thus, individuals often calculate reliability using measures of consistency. Some of the common methods for estimating reliability coefficients include test-retest, alternate or parallel forms, and measures of internal consistency (e.g., split-half, coefficient alpha). Reliability coefficients range from 0 to 1.00, with coefficients closer to 1.00 indicating higher reliability. One of the critical steps in selecting psychometrically sound instruments involves the evaluation of instruments' reliability coefficients.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Reliability coefficients are useful in evaluating instruments, and are also helpful in interpreting clients' scores and results. Standard error of measurement is another representation of an instrument's reliability, and provides a band or range of where a counselor can expect a client's "true score" to fall. Depending on the confidence level that is needed, a counselor can use standard error of measurement to predict where a score might fall 68%, 95%, or 99.5% of the time. Many professionals in the assessment field encourage counselors to interpret results using a band or range of scores based on standard error of measurement. Professionals should use standard error of difference, however, to determine whether the scores on an assessment battery are actually different.

Generalizability theory is an extension of classical test theory and involves the process of identifying different sources of error. An instrument developer should not indicate simply that an instrument has adequate reliability or that an instrument's reliability is .90. Rather, more specific information should be reported when an instrument's reliability is examined. For example, a more appropriate statement would be, "The reliability coefficient of .90 for instrument X was estimated using the test-retest method with the first administration being two weeks before the second administration. The reliability estimate was calculated using a sample of 400 high school students from two schools in rural Indiana. The demographic breakdown of the sample was as follows. . . ."

Item response theory is a family of approaches for developing instruments, which focuses on the calibration of each item within an instrument. Methods for calculating reliability/precision using item response theory are complex, but this approach is growing in popularity.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 4

Validity and Item Analysis

Learning Objectives

- L01 Define validity
- L02 Distinguish between construct underrepresentation and construct-irrelevance
- L03 Describe validation evidence based on instrument content
- L04 Explain validation evidence based on the response processes
- L05 Understand validation evidence based on internal structure
- L06 Discern validation evidence based on relations to other variables
- L07 Explain validation evidence based on consequences of testing
- L08 Distinguish between item difficulty and item discrimination
- L09 Describe the basics of item response theory

L01 Validity

Validity concerns what an instrument measures and how well it does that task. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) states: "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (p. 11). Validity is not something that an instrument "has" or "does not have"; rather, validation information informs a counselor when it is appropriate to use an instrument and what can be inferred from the results. Validity is an overall evaluation of the evidence and a determination of the degree to which evidence supports specific uses and interpretations. The interpretation of what the scores or results mean is derived from the validity evidence. For example, an instrument's title might indicate that it measures marital satisfaction; however, if we were to examine the validity information for that instrument, we might find that it is good only at predicting the number of arguments a couple has. Some people might contend that the number of arguments a couple has is not a valid indicator of marital satisfaction

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

because some couples who are deeply dissatisfied may not even talk to each other, let alone argue. It is important that helping professionals have the skills to evaluate the validity of both formal and informal instruments. In fact, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) clearly states that validity is *most fundamental* in developing and evaluating assessments.

The validation of an instrument is a gradual accumulation of evidence that provides an indication of whether or not an instrument does indeed measure what it is intended to measure. For many instruments, validity information is provided in the instrument's manual. It is not the instrument that is validated; rather, it is the *uses* of the instrument that are validated. In later chapters, we will examine instruments that are valid for certain purposes, but the validation information is not supportive of certain other usages of some instruments. Validation is a process where research is conducted to determine if the instrument's results represent the intended purpose and whether the scores or results are relevant to the proposed use (AERA, APA, & NCME, 2014). Therefore, before using an instrument, a counselor must evaluate the validation evidence and determine if there is enough support for using that instrument in that specific manner.

Reliability is a prerequisite to validity. If an instrument has too much random error, then the instrument cannot measure anything consistently. The reverse can also be true, however, where an instrument is reliable but not valid. For example, an instrument may measure something consistently, but it may not be measuring what it is designed to measure. Therefore, although reliability is necessary for an instrument to be sound, high reliability does not guarantee that the instrument is a good measure. In selecting assessment tools, a counselor should first examine the reliability and see if the instrument measures consistently. The counselor should then move to *what* the instrument measures and *how well* it measures what it was designed to assess. It is inappropriate to say that an instrument has validity because the validation evidence will only provide support for specific uses.

L02

In order to understand validation evidence, it is important to understand the terms **construct underrepresentation** and **construct-irrelevance**. In assessment, the term *construct* means the concept or characteristic that the instrument is designed to measure (AERA, APA, & NCME, 2014). When considering validation evidence, the counselor should see a compelling argument for using the instrument in specific ways, which includes proposition or claims for appropriate uses. The counselor needs to be a good consumer of this evidence, which involves some skepticism and considering rival hypotheses. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), plausible rival hypotheses can often be fashioned by considering whether the instrument measures less or more of what it is proposed to measure. Construct underrepresentation concerns the degree to which the instrument is unable to capture significant aspects of the construct. For example, if an instrument is designed to measure problems related to eating and only measures attitudes toward food and does not measure any behaviors related to eating (e.g., binging). On the other hand, construct-irrelevance refers to the degree to which scores or results are affected by factors that are extraneous to instrument's intended purpose. This occurs when the scores are systematically influenced by processes that are not part of the construct. An example of this would be a mathematics test that relies heavily on word problems that may be assessing reading level to some extent. Construct underrepresentation and construct-irrelevance should both be considered when examining validation evidence.

In the field of assessment, there has been a movement away from the three traditional categories of validity (i.e., content-related, criterion-related, and construct-related evidence). Historically, **content-related validity** concerned the degree to which the evidence indicated that the items, questions, or tasks adequately represent the intended behavior domain. For example, content validity traditionally involved examining the content

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

or items, such as does a test of third-grade spelling contain typical words that educators would expect third graders to know how to spell. **Criterion-related validity** concerned the extent to which an instrument was systematically related to an outcome criterion. With criterion-related validity, we were interested in the degree to which an instrument was a good predictor of a certain criterion. An example of criterion-related validity is whether the SAT predicts academic performance in college. The third historical type of validation evidence is **construct validity**, which concerns the extent to which the instrument may measure a theoretical or hypothetical construct or trait. Although readers should be aware of the three historical types of validity (i.e., content-related, criterion-related, and construct), many experts in measurement now argue that separating validation evidence into these three distinct types obscures the notion that validity is a unitary concept in which there is a gradual accumulation of evidence that supports the instrument's specific use.

This view of validation as a unitary concept is shared by a number of leaders in the assessment field, such as Messick (1995), who contended that dividing validity into three types only fragments the concept. Messick argued that a comprehensive view of validity integrates consideration of content, criteria, and consequences into a construct framework in which the practitioner then evaluates the empirical evidence to determine if it supports the rational hypotheses and the utility of the instrument. Messick suggested that validity is best viewed as a comprehensive view of construct validity. Messick further argued that a more comprehensive view of validity should take into account both score *meaning* and the *social values* in test interpretation and test use. Therefore, in evaluating the empirical evidence, the counselor needs to consider both score meaning and the consequences of the measurement.

The gathering of validation information is analogous to the techniques used in the scientific method where the starting point is the instrument and theories about the trait or construct. Predictions are made related to the instrument based on theory or previous research. These predictions are empirically tested, and the results of those empirical tests are either supportive of the validity of the instrument or not. This process is continually repeated, and the accumulation of evidence is evaluated to determine an instrument's validity. Counselors need to review the accumulation of information and determine the appropriate uses of the instrument.

Table 4.1 includes different methods for acquiring validation evidence according to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). It should be noted that validation is a shared responsibility among the developers and the practitioner. The instrument's developer is responsible for providing relevant research and

TABLE 4.1
Validation evidence from the Standards for Educational and Psychological Testing

- **Evidence Based on Test Content** Concerns analyzing the evidence of the relationship between the instrument's content and the construct it is intended to measure.
- **Evidence Based on Response Processes** Includes theoretical and empirical investigations of individual's cognitive, affective, or behavioral response processes. (This evidence concerns whether the examinee is responding in a manner consistent with the intent of the assessment.)
- **Evidence Based on Internal Structure** Involves analyzing the internal structure of the instrument to determine whether it conforms to the intended construction of the instrument. For example, factor analyses could be performed to examine whether the interrelationship of items corresponds to the appropriate subscales of an instrument.
- **Evidence Based on Relations to Other Variables** Entails examining the relationship between the assessment scores and other pertinent variables. For example, investigating whether an assessment of depression correlates with other measures of depression.
- **Evidence for Validity and Consequences of Testing** Concerns the evidence about the intended and unintended consequences of the use of this instrument. For example, studies of whether an employment test results in different hiring rates for different groups of people.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

evidence related to using the instruments and the counselor is responsible for determining whether sufficient evidence supports using it in assessing a certain client or clients. Furthermore, frequently other researchers and not just the instrument developer conduct research that adds to the validation evidence. The type of instrument and how the instrument is going to be used will influence the methods used to gather validation evidence. In some circumstances, the evidence related to items or content of the instrument may be particularly pertinent. In other circumstances, a counselor would be particularly interested in evidence concerning whether the instrument predicts certain behaviors (e.g., does the instrument predict clients who are likely to attempt suicide?). These different types of validation evidence should not be considered as distinct types of validity because validation requires examining the bulk of evidence and determining whether the assessment is appropriate for certain uses.

L03 Evidence Based on Instrument Content

Although the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) uses the term *test* (see Table 4.1), I am going to use the term *instrument* in discussing methods related to gathering validation evidence. Evidence based on instrument (or test) content is related to the traditional term *content validity.* Often practitioners need to examine the evidence related to the development of items, questions, or tasks to consider whether the content of the instrument seems to adequately represent the intended construct. As an example, let's pretend (although it may not be merely a fantasy for some readers) that you are going to take a test soon covering basic principles of assessment. You may carefully study the content in the first four chapters of this book and feel quite prepared to take the test. You take the test, however, and find that all of the questions address concepts solely related to reliability! You could then argue that there was a problem with the content of the test because, with all the questions coming from one chapter, the content domain (basic assessment concepts) was not adequately sampled or represented. In providing validation evidence related to the content of an instrument, the developers should discuss how they analyzed the construct and how they selected the items or questions that measured the construct.

With evidence related to the content of an instrument, the central focus is often on how the instrument's content was determined. Documenting the procedures used in instrument development frequently provides validation evidence. Instrument developers should first begin by identifying the behavior domain they wish to assess. At first glance, this seems to be an easy step, but in actuality, it is often one of the more difficult ones. In counseling, there are often different theoretical perspectives concerning the areas we wish to assess (e.g., self-esteem, personality, career interests), and there is no universal agreement on the definitions of these constructs. Let's use an example of developing an instrument to measure client change that may occur because of counseling. Consistent with other counseling topics, there are various views on what constitutes client change and how to measure it. Hence, if we are attempting to construct a measure of client change, we would first need to define and conceptually clarify this concept. After initially clarifying the concept, we would then draw up *test specifications.* Test specifications provide the organizational framework for the development of the instrument. A common first step in test specification is to identify the goals or the content areas to be covered. In the example of client change assessment, we need to bear in mind that client change is a very broad area. Therefore, the instrument developers would need to consider whether to measure behavioral change, affective change, or cognitive change. The process of identifying goals and content is helpful because it requires us to clearly articulate the intended purpose(s) of the instrument. The instrument's manual should include a rationale for each intended use of the instrument and provide theoretical and empirical information relevant to those intended usages.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The next step in test specification is to precisely state the objectives or processes to be measured. For example, in client change, who will assess the change (e.g., client report, counselor report, or someone else)? Once the objectives have been established, they need to be analyzed, and the relative importance of each objective needs to be considered. We would probably want to include more items related to objectives that are of greater importance. Before writing any of the items, the test specification process also involves determining the level of difficulty or level of abstraction of the items. For example, in the assessment designed to measure client change, we would need to determine whether the instrument was going to be for children or adults, the level of complexity of the items, and the appropriate reading level. Any instrument manuals should describe in detail the procedures used in developing the instrument before any items were written. Developers create a more refined and better quality instrument if significant work is exerted before the items are written.

Once the items are written, another method of providing content-oriented evidence involves having experts review the instrument. Experts can be scholars in the instrument's content area or, perhaps, practitioners working in the field. Experts will typically analyze the degree to which the instrument's content reflects the domain, whether the weighting of the instrument is appropriate, and the degree to which the content is assessed in a nonbiased manner. In general, these critiques generally will lead to revisions of the instrument. The manual should provide the number of experts used, a description of the experts, and an explanation of why these individuals were selected. If experts' judgments were used, then the manual should describe what was done to respond to these judgments.

Content-oriented validation evidence should not be confused with *face validity*, which means what it says—on the surface, the instrument looks good. Face validity is not truly an indicator of validity and should not be considered as one. An instrument can look valid but may not measure what it is intended to assess. Remember that an instrument is not valid or invalid; rather, the validation evidence informs the clinician of the appropriate uses of that assessment. Therefore, a helping professional cannot simply examine an instrument's questions or items and decide to use that assessment tool. Counselors also need to evaluate the content-oriented validation evidence provided in the instrument's manual and determine if there is sufficient evidence to use it. Although validation evidence related to content is always important, there are situations in which it is particularly vital. An example is with achievement tests, where we are interested in how much an individual has learned. We cannot expect someone to learn something unless he has had exposure to that information. Sometimes this concept is called *alignment* and it concerns the level of correspondence between student learning standards and the content of the test. Therefore, there needs to be a direct correspondence between the student learning standards and the contents of the assessment or there are problems with the assessment. Content-sampling issues can become quite important as the assessment may not adequately represent the curriculum standards and this sampling procedure may result in construct underrepresentation or construct-irrelevance.

LO4 Evidence Based on the Response Processes

As researchers have learned more about information processing, this knowledge has been applied to instrument development. Analyzing the cognitive processing of clients while they are taking the assessment is important in some assessment situations. The goal of this type of validation evidence concerns whether individuals either perform or respond in a manner that corresponds to the construct being measured. For example, if an assessment is attempting to measure vocational interest, then the instrument developers might examine whether people are answering the items based on what they like to do rather than what

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

they are able to do. An example of how researchers might examine this is to have individuals "think aloud" while taking the interest inventory. The test takers' verbalizations of their thoughts while taking the interest inventory would be recorded and researchers then analyze the manner in which the test takers processed the inventory to see if they were considering their interests while taking the assessment.

Another example of gathering validation evidence based on response processes is to question examinees about their performance strategies or responses to items. As an example, if it is an assessment of mathematical reasoning, the examinees would be queried about their strategies for solving each problem to see if they used the intended mathematical concepts. Sometimes evidence can be gained by analyzing factors such as eye movements or response times.

Sometimes it is interesting to examine if there are subgroup differences (e.g., race or ethnicity) regarding the processing of information. It may be that a distractor within an item confuses individuals from certain groups when they try to apply the intended mathematical concepts, but does not distract members from other groups. By examining the response process used by different groups, the test developers can rewrite the item so that it is processed in the same manner by individuals from diverse groups.

Particularly in the area of achievement, there have been significant advancements concerning models or theories of cognition and learning. Based on these models and theories, there can be computer-simulated performances on assessment items and then researchers can compare the performance of children or adults to the model-based answers produced by computer. Hence, if individuals' responses match the computer's theory-developed answer, then there is some evidence that individuals are using the intended theory. Validation evidence based on response processes can often provide keen insights about an instrument, and instrument manuals are increasingly including this type of evidence.

LO5 Evidence Based on Internal Structure

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) suggests that analyses of the internal structure of an instrument can also contribute to the validation evidence. An instrument may be designed to measure one clearly defined construct or it may have subtests or scales within the instrument. In analyzing the validation evidence, it is often useful to examine whether the items in one subscale are related to other items on that subscale, a task that involves examining the internal structure of the instrument. For example, the NEO Personality Inventory—3 (NEO-PI-3) is designed to measure five factors of personality: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness (McCrae & Costa, 2010). When using the NEO-PI-3, a counselor should examine whether the items that measure neuroticism are related because it would be disconcerting if several of the items designed to measure neuroticism were more closely related to items on the agreeableness subscale as compared to items on the neuroticism subscale.

Factor analysis is often used to examine the internal structure of instruments because factor analysis is a statistical technique used to analyze the interrelationships of a set or sets of data. There are a variety of reasons for using factor analysis in gathering validation evidence, including exploring patterns of interrelationships among variables. Hence, results of a factor analysis provide information on the pattern of interrelationships among items and a clinician can see if those interrelationships correspond to the appropriate subscales. In examining the internal structure of the instrument, factor analysis can also identify items that may be redundant (i.e., very highly related to each) so that instrument developers may present results in a manual to explain their process for reducing the number of items in a subscale or in an instrument. There are various forms of factor analysis including

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

exploratory and confirmatory methods. As the names imply, exploratory methods for factor analysis explore the factor structure of the instrument; confirmatory methods use the existing structure of the instruments and examine whether the factor analysis matches or confirms that structure.

Also, as will be discussed in Chapter 6, researchers study the internal structure of instruments with different groups, such as men and women or individuals from different racial or ethnic backgrounds. Some studies regarding the internal structure of an instrument involve examining whether there are group differences on individual items, which is typically called **differential item functioning**. As will be discussed in more detail in Chapter 6, differential item functioning examines whether there are group differences (e.g., Caucasian, African American, Latino(a)) in terms of the percentage of individuals who get the item correct. If, for example, 95% of Caucasians get an item correct but only 50% of African Americans of similar ability get that item right, there may be something within the item that results in this differential functioning. These results indicate that the item should be examined. Another method for examining the internal structure and group difference involves factor analysis and subscales. It sometimes happens that results from factor analyses indicate that the internal structure for an instrument corresponds to the intended design of the instrument for one group but the internal structure of an instrument might be quite different for another group.

LO6 Evidence Based on Relations to Other Variables

Another commonly used method for gathering validation evidence is to analyze the instrument's relationship with other variables. For example, an instrument developer may want to show that an instrument designed to measure depression is associated with other indicators of depression (e.g., feeling blue, lack of interests). When the validation evidence concerns examining the relationship of the instruments with other variables, clinicians need to consider the quality of the other variables. An instrument may be highly related to another variable, but if the other variable is not anything meaningful, then that greatly reduces the likelihood that you would want to use that instrument with clients. In examining the validation evidence, counselors want to see that the instrument is related to other variables that are *relevant* and *pertinent* to client issues or that add to our knowledge about using that instrument.

A statistical tool often used in providing validation evidence related to an instrument's relationship with other variables is the **correlational method**. As discussed in Chapter 3, the statistical technique of correlation is used to explore the relationship between two variables. Counselors often want to examine the degree to which an instrument is related to other pertinent variables. Because the focus is on the degree to which there is a relationship, it is logical to use correlation to provide an indication of the magnitude of that relationship. There are several steps in gathering validation evidence using correlation. First, select an appropriate group to use in the validation study. Second, administer the instrument. Depending on how the instrument is going to be used, sometimes the criterion data is gathered immediately or with instruments that predict future behavior—there may be a time lapse before the information is gathered (e.g., waiting until the group finishes college to gather academic performance information). Once the information on the other variable is collected, the final step is to correlate the performance on the instrument with the information from the other variable. The result of that calculation is often labeled a **validity coefficient**.

Validity coefficients can be helpful to counseling practitioners in two ways. Counselors can compare the validity coefficients from different instruments and select the instrument that has the highest relationship between the instrument and the relevant variables. Although this sounds quite easy, in practice it can sometimes be complex because there are

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

correlations with a number of different variables. The second way that validity coefficients can be useful in practice is by allowing the counselor to examine the amount of shared variance between the instrument and pertinent variables. As you will remember from the discussion of correlation in Chapter 3, a correlation coefficient can be squared, resulting in a *coefficient of determination*, which provides an indication of the amount of shared variance between the two variables. Thus, by squaring the validity coefficient, we get the percentage of shared variance between the assessment instrument and the other relevant variables. In many discussions of validity coefficients, a question arises about the magnitude of the coefficient and how large a validity coefficient should be. Kaplan and Saccuzzo (2012) indicated that validity coefficients are rarely larger than .60 and that they often fall in the .30 to .40 range. Anastasi and Urbina (1997) suggested that validity coefficients should at least be statistically significant, which means we can conclude that the coefficient had a low probability of occurring by chance.

As indicated earlier, the constructs of interest in counseling (e.g., depression, academic achievement, work values) are complex and, therefore, evidence of validity cannot be confirmed with one correlation coefficient. Consequently, when examining an instrument's manual, counselors will often find that the validation of an instrument has been explored by seeing if it relates to other variables that are consistent with its use and seeing if the instrument has a low relationship with incompatible constructs. Examining an instrument's relationships with similar constructs and purportedly different variables is called convergent and discriminant evidence. **Convergent evidence** means an instrument is related to other variables to which it theoretically should be positively related. For example, if an instrument is designed to measure depression and correlates highly with another instrument that measures depression, then there is convergent evidence. **Discriminant evidence**, on the other hand, exists when the instrument is not correlated with variables from which it should differ. For example, there is evidence of discriminant validity when an instrument designed to measure depression is *not* significantly correlated with an instrument that measures trait anxiety. According to Anastasi and Urbina (1997), discriminant validation information is particularly important in terms of evaluating the validity of personality instruments, in which irrelevant constructs may influence scores in assorted ways.

Campbell and Fiske (1959) suggested using a **multitrait-multimethod matrix** in order to explore validation evidence. This procedure involves correlating an instrument with traits that are both theoretically related to the instrument and traits that should be unrelated to the construct being measured. This process also involves correlating the instruments with other instruments that use the same and different assessment methods (e.g., self-reports, projective techniques, behavioral ratings). The result of these various correlations is a matrix that enables professionals to evaluate the relationship between the instrument and measures of (a) the same traits using the same testing method, (b) different traits using the same testing method, (c) the same traits using different testing methods, and (d) different traits using different testing methods. Support for using the instrument would be indicated when the correlations are higher between the same trait using different assessment methods as compared with different traits using the same methods or with different traits using different methods.

It should be remembered that correlation coefficients provide only an indication of the relationship between the instrument and another variable. Therefore, if the intent is to *predict* behaviors based on the instrument's results, then researchers need to use additional techniques to see if the instrument can be used for prediction.

Prediction or Instrument-Criterion Relationships. Sometimes we want an instrument that predicts behavior and, therefore, the validation evidence should focus on the degree to which the instrument does predict relevant behaviors. In assessment, when an

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

instrument was developed to predict or identify, then traditionally users would examine the **criterion-related validity**. The criterion variable is what the instrument is designed to predict and this variable is distinct from the instrument. Examples of criterion variables are academic performance in a graduate program, which is predicted by the GRE or job performance or performance in a training program, which is predicted by the Armed Services Vocational Aptitude Battery (ASVAB).

Historically there are two types of validation evidence related to prediction: **concurrent validity** and **predictive validity**, with the difference being the period of time between taking the instrument and gathering the criterion information. In concurrent validation, there is no time lag between when the instrument is given and when the criterion information is gathered. Prediction is used with concurrent validity in a broad sense because it is predicting behavior in the current context. This type of criterion-related validity is used when we want to make an immediate prediction, such as with diagnosis. On the other hand, with predictive validity, there is a lag between when the instrument is administered and the time the criterion information is gathered. For example, if we wanted to develop an instrument that would identify couples who will stay married for more than 10 years, we would administer the test right before a couple was married and then wait 10 years to gather the criterion evidence, which would be whether the couple was still married.

In considering variables that an assessment is designed to predict (i.e., the criterion), we should be particularly interested in precisely what the criterion is and its psychometric qualities. A good criterion is truly a variable of interest. As an example, there would be problems if I were developing an instrument to measure resiliency and used as my criterion whether I simply thought the person looked resilient. The criterion should be *reliable* and relatively free from random error. If there is a large degree of error in the criterion, then there is a very good possibility that the instrument will be unable to predict anything useful. Using an assessment without a reliable criterion is like trying to predict to a target that is always moving. We also want the other variable to be *free from bias.* As an illustration, if an instrument were developed to predict a firefighter's effectiveness, the criterion should be an assessment of job performance. It would be a problem if the criterion variable was supervisors' evaluation of job performance, and the supervisors used in this validation study happened to be biased against female firefighters. Finally, we want the other variable to be *immune from criterion contamination.* Criterion contamination occurs when knowledge of the instrument results or scores influences the gathering of criterion information. For example, criterion contamination may occur if those determining a *DSM-5* diagnosis had prior knowledge of a participant's performance on an instrument designed to diagnose. A practitioner could easily be influenced by the information from the instrument and might tend to assign a diagnostic label that was consistent with the instrument. In this example, the correlation between the instrument and the diagnostic categories would appear artificially high. Thus, the criterion, in this case, is "contaminated" by the prior knowledge of the individual's performance on the diagnostic instrument.

In providing validation evidence related to an instrument's abilities to predict behavior, researchers often use regression-related statistical techniques or expectancy tables.

Regression. The statistical technique of **regression** is another commonly used method for gathering validation evidence. Regression, which is closely related to correlation, is commonly used to determine the usefulness of a variable, or a *set* of variables, in predicting another important or meaningful variable. In assessment, we are usually specifically interested in whether the instrument predicts some other specific behavior (the criterion variable). For example, does an instrument designed to measure potential suicidal behaviors actually predict suicidal behaviors? Regression is based on the premise that a straight line,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

called the *regression line*, can describe the relationship between the instrument's scores and the criterion. In Figure 4.1, the scores on an instrument are plotted in relation to the scores on the criterion. The line that best fits the points is the regression line. Once a regression line, or line of best fit, is established, it is possible to use it to predict performance on the criterion based on scores on the instrument.

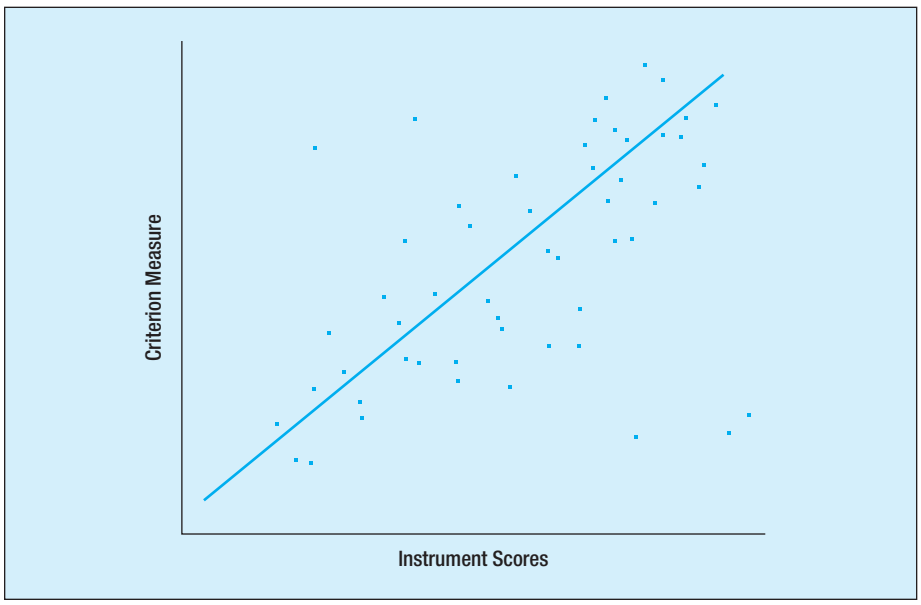
To illustrate the concepts of regression, imagine you are interested in an instrument that predicts people's abilities to be "fun" (please note that this is only an illustration). In this example, you are comparing three instruments to use for the screening of friends, neighbors, and coworkers. Figure 4.2 shows three scatterplots that describe the relationship among three instruments and the criterion. In all three of these cases, an individual took one of the three instruments and then friends of that individual evaluated the person's ability to be fun. Thus, the criterion, which is plotted on the *y*-axis, is a rating of "funness." By examining the diagrams in Figure 4.2, we can see that Instrument A is a pretty good predictor because all scores are located near the line. For example, most of the individuals who received a score of 50 were rated as being around 3.00 in terms of funness. Instrument B, however, is not quite as good a predictor because the points are more spread out. With Instrument B, there is wide variation for those individuals who received a score of 50, with the evaluation of funness ranging from 2.00 to 4.00. There is such a poor relationship between Instrument C and the funness ratings (the criterion) that a regression line or a line of best fit could not be determined. To help in making certain social decisions, you may want more precision in your prediction of "funness" than either Instrument B or Instrument C can provide.

Rarely in assessment does a scatterplot provide evidence of an instrument's predictive ability; rather, the evidence is provided by using the statistical technique of regression, which tells us if we can use a **regression equation**. This equation describes the linear relationship between the predictor variable and the criterion variable. The regression equation is

$$Y^1 = a + bX$$

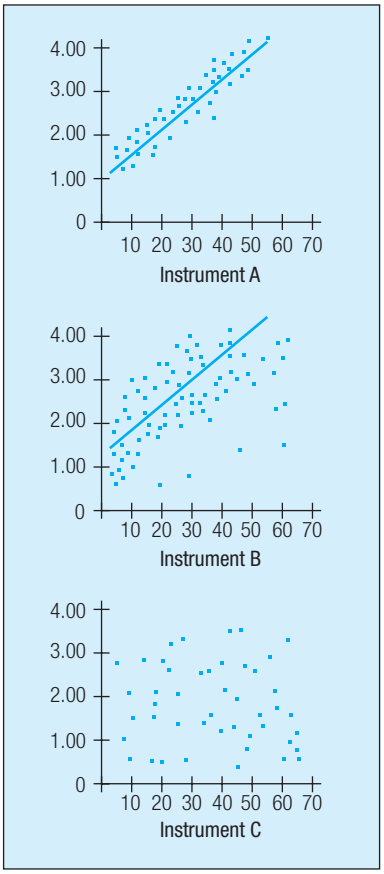
where Y^1 is the predicted score; a is the *y-intercept*, or the *intercept constant*; b represents the *slope*, which is the rate of change in Y as a function of changes in X; and X is the

FIGURE 4.1
Regression line



Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 4.2
Scatterplots between scores on three instruments and the criteria for these instruments



individual's score on the instrument. The slope is usually called the *regression coefficient* or the *regression weight.* Thus, in an instrument's manual, the instrument developers will provide information on whether the instrument can significantly predict the criterion (i.e., Can the data be fitted into a significant regression equation?). While on the subject of regression, note that there is also *multiple regression,* in which there are multiple predictor variables. As an example of multiple regression, suppose that two predictor variables, a precollege test (e.g., SAT or ACT) and high school grades, are used to predict academic achievement in college. In multiple regression, the equation is extended to include the additional variables that have different weightings ($Y^1 = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_nX_n$). In using multiple regression, it is possible to determine which variables are the best predictors as well as how much variance is explained when additional variables are added to the equation. Hence, it is possible to analyze whether high school grades or a precollege test is a better predictor of college performance and the amount of variance that these variables explain, both in combination and separately.

In counseling, just as no instrument can have perfect reliability, no instrument is perfect at prediction. As discussed in Chapter 3, we take the lack of perfect reliability into consideration and provide scores in a range using standard error of measurement. The same logic applies with predicting to a criterion, where the lack of a perfect relationship between the instrument and the criterion is taken into consideration in the interpretation of the criterion scores by using **standard error of estimate**. Standard error of estimate indicates the margin of expected error in an individual's predicted criterion score as a result of imperfect validity. The formula for standard error of estimate (SE_{est}) is similar to the formula for

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

standard error of measurement except that the focus of standard error of estimate is on the validity coefficient:

$$SE_{est} = S_y\sqrt{1 - r_{xy}^2}$$

In this equation, s_y is the standard deviation of the criterion and r_{xy} is the validity coefficient (correlation between the scores on the instrument and the criterion), which is squared (r^2_{xy}). Although standard error of estimate can be used for predicting an individual's performance on the criterion, it is more commonly used to look at error around a cutoff point. A developer of an instrument designed to predict suicidal behaviors would want to use standard error of estimate when determining the score at which counselors should be concerned about their clients. As with standard error of measurement, standard error of estimate provides a range of scores; however, with standard error of estimate, the range is on the criterion.

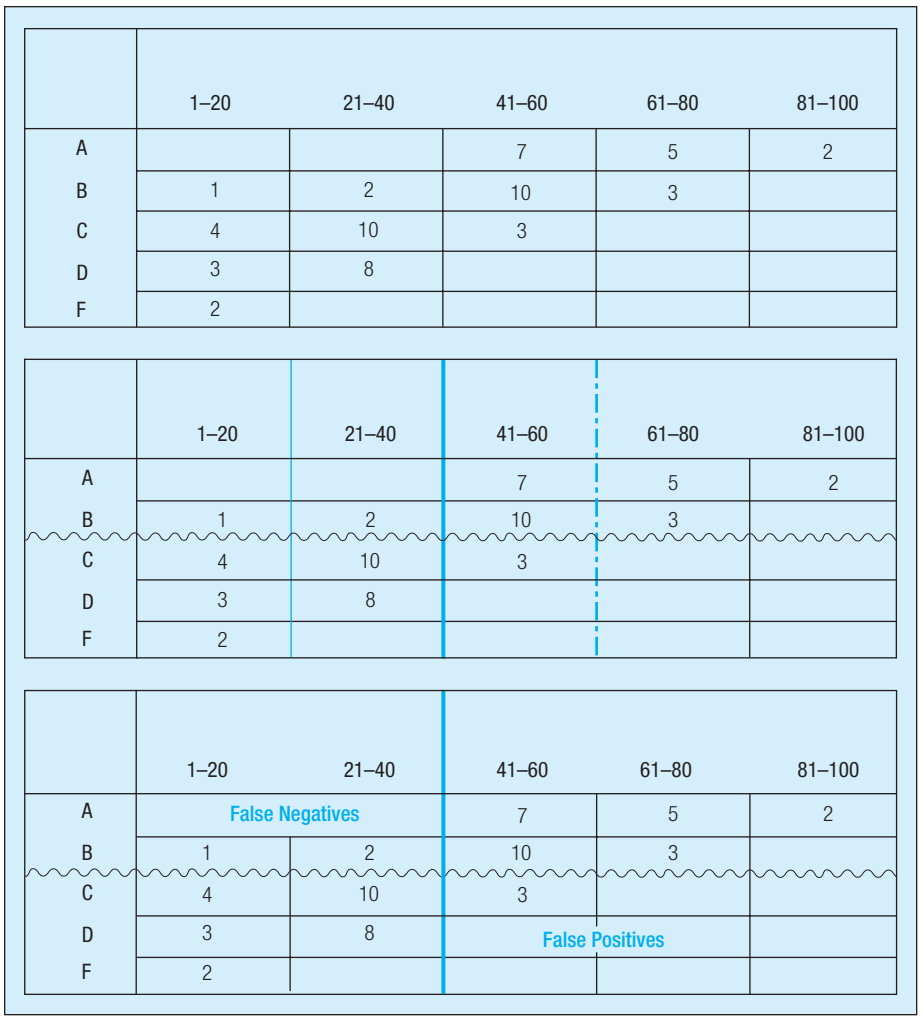
Decision Theory. Another method that examines the relationship between an instrument and a criterion or predictor variable is *group separation* or *expectancy tables.* This approach is sometimes referred to as **decision theory** because instruments are often used in decision making, and this approach evaluates the accuracy of the decision making. Group separation concerns whether the scores of the instrument correctly differentiate. For example, do people who score high on the test get the high grades, as predicted? Decision theory attempts to assist in the selection and placement process by taking available information and putting it into a mathematical form (see Cronbach & Gleser, 1965). In terms of aiding the decision-making process, we want to know how often the instrument is right (a *hit*) and how often the instrument is wrong (a *miss*). For example, to what extent does the instrument accurately predict those who will succeed in college versus those who will not succeed? An **expectancy table** is the table that is developed to examine whether an instrument differentiates between the groups. An expectancy table can be used in two ways. First, an expectancy table can help in making decisions, such as admissions or hiring decisions. Second, it can provide clients with information on the probabilities of succeeding in college or an occupation given their scores on the instrument.

Let's return again to the fictitious Counseling Aptitude Scale (CAS). As you will remember, the CAS is supposed to predict who will be effective counselors and who will not. If we were the developers of this instrument, we could use an expectancy table to see (a) if the CAS does predict effective counseling and (b) if we could use the CAS to determine cutoff scores for admission to counseling programs. After we develop the expectancy table, we could then use it to provide potential students with an indication of how successful they might be as counselors. To develop an expectancy table, we must first find a criterion to use in validating the instrument.

Hypothetically, the CAS will predict counseling effectiveness. Because this is a difficult concept to measure, we might want to use performance in a graduate program in counseling as a criterion. (I know some will argue that performance in a graduate program is not a very good measure of people's effectiveness in actual counseling, but let's just go ahead with this as an example.) After we have established the criterion of grade point average (GPA) in a counseling program, we next have to gather data. To do so, we would take a group of students who had been admitted to a counseling program, and, before they began classes, we would give them the CAS. We would then sit back and let these students proceed with their counseling program. After the students graduated, we would then gather GPA information to construct our expectancy table. Figure 4.3 shows an imaginary expectancy table for the CAS scores and the GPAs of 60 students. We would examine the first table to determine if the CAS can differentiate between those students who did well in a graduate counseling program and those who did not perform well.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 4.3
Expectancy tables for the Counseling Aptitude Scale



	1–20	21–40	41–60	61–80	81–100
A			7	5	2
B	1	2	10	3	
C	4	10	3		
D	3	8			
F	2				

	1–20	21–40	41–60	61–80	81–100
A			7	5	2
B	1	2	10	3	
C	4	10	3		
D	3	8			
F	2				

	1–20	21–40	41–60	61–80	81–100
A	False Negatives		7	5	2
B	1	2	10	3	
C	4	10	3		
D	3	8	False Positives		
F	2				

In examining this table, we can see that those who did well on the CAS tended to have higher GPAs than those who did poorly on it. We would then examine this table to determine how many times the instrument "hit" (was an accurate predictor) and how many times it "missed" (was not an accurate predictor). Examining the hits and misses not only helps to evaluate the criterion-related validity of the CAS; it can also aid us in determining cutoff scores for use in future admissions decisions.

Often in graduate school, a student must earn a GPA of B or above to continue in a program. Therefore, in the second table in Figure 4.3, a squiggly line has been drawn under the B row to indicate which students earned a B or better. As you can see from the second table, 30 students had GPAs above B (1 with scores of 1–20, 2 with scores of 21–40, 17 with scores of 41–60, 8 with scores of 61–80, and 2 scores of 81–100). If we set the cutoff score at 21 and above (signified by the single line), then we would have 38 hits and 22 misses. The 22 misses would include the one person who received a score between 1 and 20 but achieved a GPA of B even though that was not predicted, as well as the 18 who scored between 21 and 40 but had grades of either C or D and the 3 people who scored between 41 and 60 who received a C. What we now need to do is to determine the hits and misses if the cutoff score is set at 41 and above (signified by the dark line) and 61 and above (signified

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

by the line with dashes). If we set the cutoff score at 41, there are 54 hits and only 6 misses. On the other hand, if we set the cutoff score at 61, there would be 40 hits and 20 misses. Therefore, if the decision were based solely on the most hits and the least misses, then the cutoff score of 41 seems best.

In addition to identifying hits and misses, expectancy tables can provide information about **false positives** and **false negatives**. A false positive occurs when the instrument predicts that individuals have "it" (the criterion) when in fact they do not. Thus, if we examine the third table in Figure 4.3, we can see three false positives—the three individuals who had scores at or above the 41-point cutoff but who received a C average in graduate school. A false negative occurs when the instrument predicts that test takers do not have it when in fact they do. Figure 4.3 shows three false negatives—individuals who had a score of less than 41 but earned a B average in graduate school. Hence, the assessment predicted these students would not do well, but the prediction was a false negative because the students actually had a B average.

The assessment situation influences whether individuals should try to minimize the false positives or the false negatives in setting cutoff scores. For example, if an instrument were used to select individuals for a very expensive training program, we might want to minimize the number of false positives. The purpose would be to select only people who would be successful and to not spend the money to train people who wouldn't. On the other hand, a false negative can also be quite costly. For example, a false negative on an instrument designed to assess suicide potential means the instrument would not identify someone as being at risk for committing suicide when, in fact, he or she is at risk. In this case, a false negative could mean the loss of life. Sometimes expectancy tables are used to identify the false positives or false negatives and then identify other characteristics of those individuals (e.g., personal characteristics, background characteristics) that may contribute to the misclassification. Using the last table in Figure 4.3, we might find it interesting to gather more information about the three individuals who scored between 41 and 60 but who did not achieve a B average in graduate school. There may be characteristics about these individuals that could help in screening those who appear to have counseling potential based on their CAS scores but who do not do well in their graduate counseling program.

Expectancy tables are commonly used for selection and placement purposes. They can also be used in counseling to provide probability information to individuals (e.g., What is the probability of achieving an A average given a CAS score of 35?). The probability information is calculated by determining the percentage of people who fall in certain criterion categories. In looking at Figure 4.3, for example, we see that the probability of achieving an A average with a CAS score of 35 is zero because no one who received a score of 21–40 achieved an A average. There is, however, a 10% chance that a person receiving a score of 35 could achieve a B average because 2 of the 20 people who scored between 21 and 40 received a B average. The probability of receiving an A average with a CAS score of 90 is 100% because all of the people who scored between 81 and 100 achieved an A average.

Other factors need to be considered when using expectancy tables to determine cutoff scores for selection or placement purposes. One factor concerns the degree to which the process is *selective.* In many ways, a cutoff score is easier to determine if the process is very selective. The opposite is also true—the selection process is simplified if the intent is to select most individuals and only weed out a few who are extremely low. Another factor that influences the decision-making process is *base rate,* which refers to the extent to which the characteristic or attribute exists in the population. The third factor that influences the decision-making process is the way in which the scores *distribute.* The selection process is considerably more difficult when scores are grouped closely together than when there is more variability in scoring.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Validity Generalization. In the *Standards for Educational and Psychological Testing* (APA, AERA, & NCME, 2014), the last subcategory related to validation evidence based on relations with other variable is *validity generalization.* Validity generalization was initially debated in employment testing and concerned the degree to which test validation research could be generalized to new situations. In general, findings from one study are not conclusive and there is a need for replication. When there are numerous studies related to an instrument, then there are methods for analyzing the data to determine if the relationship among the instrument and criterion are related to statistical artifacts or research methodologies or if the relationship is quite robust and can be generalized. The statistical procedure used to examine whether an instrument's validity can be generalized is *meta-analysis.* In meta-analysis, researchers combine the results of many studies and examine the influences of different research factors to determine whether a particular instrument has a significant relationship with a criterion.

Generalizing the validation evidence requires that a large number of studies have been conducted with the instrument. Furthermore, the researchers must employ complex meta-analytic procedures that allow for the examination of consistent patterns of relationships between the instrument and criteria. After this type of complex meta-analytic research is completed, researchers examine the findings to determine if the results warrant a generalization of the previous validation evidence to other uses of the instrument. An example of validity generalization is related to research conducted by Hunter and Schmidt (1983) and their colleagues regarding general ability or intelligence scores and job performance. Their research is discussed in Chapter 9.

LO7 Evidence Based on Consequences of Testing

Another topic related to validity evidence that the *Standards for Educational and Psychological Testing (*AERA, APA, & NCME, 2014) clarifies is the concept that practitioners have a responsibility to examine the evidence related to the consequences of testing when considering using an instrument. Often in counseling, there is an expectation that either the results of the instrument will be helpful to the client or the interpretation of the results will produce positive outcomes for clients. The validation evidence should also be examined to determine if these positive results are actually realized.

In examining the evidence of consequences of testing, a counselor should also consider the social consequences of the assessment. Examples of social consequences of assessment are group differences in test scores on instruments used for employment selection and promotion and group differences in the placement rate of children in special education programs. Validity is a unifying concept in which the counselors simultaneously examine the evidence considering use, interpretation, and social implications. Considering both validation evidence and issues of social policy becomes particularly important in cases when there are differing consequences related to test usage among different groups. Practitioners should evaluate the evidence to determine if there is a compelling argument that justifies the test use and the appropriate interpretation. This concept of considering social implications is particularly relevant to counselors who have the ethical responsibility to consider what is in the client's best interest when they are counseling clients.

In some situations, it is suggested that testing will result in some defined benefits, such as increased student motivation. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) suggests that these benefits should not be automatically accepted and in evaluating the validation evidence, clinicians need to examine the research related to the intended and unintended consequences of such testing. As argued earlier, counselors are responsible for conducting assessments that are in the best interest of the client and, therefore, it is a counselor's responsibility to investigate the evidence or lack of evidence regarding the benefits or deleterious effects of testing.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Conclusion on Validation Evidence

In conclusion, the validity of an instrument is indicated by the gradual accumulation of evidence; consequently, there is not a clear-cut decision on whether an instrument has validity or not. It must be remembered that validation evidence provides the basis for our interpretation of the results. To interpret the results of an instrument appropriately, a counselor must evaluate the validity information and determine what can be inferred from the results. Cronbach (1990) suggested that the validation evidence should be seen as a persuasive argument in which we are convinced by the preponderance of evidence on the appropriate use of an instrument. An instrument with strong validation evidence will provide confirmation that critical factors and rival hypotheses have been examined. Many times the evaluation of the validation evidence need not be done by the counselor in isolation, because many instruments have been critiqued by knowledgeable scholars in publications such as the *Mental Measurements Yearbooks.* It should be noted, however, that the validation process for an instrument never ends as researchers are frequently studying assessment instruments; therefore, clinicians will need to stay abreast of current research. Furthermore, the extent to which there needs to be validation evidence will vary depending on the use of the instrument. For example, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) indicates that "higher stakes may entail higher standards of evidence" (p. 22).

The emphasis in this chapter has been, to a degree, on formal assessment tools; yet, the concepts associated with validity apply to *all* types of client assessment. For example, counselors need to consider the concepts of validity when they do an intake interview with a client. In interviewing a client, a counselor needs to consider the content of the interview and the evidence that the interview questions address the appropriate behavior domain. Furthermore, counselors often make predictions about clients' behaviors and they should examine what informal assessments they are basing those predictions on and whether there is an accumulation of evidence that supports those hypotheses. Furthermore, counselors need to consider the clinical decisions they are making about clients based on informal assessments and the evidence that supports these clinical decisions. Research tends to support the contention that practitioners who rely only on intuition and experience are biased in their decisions (Dawes, 1994). Therefore, it may be helpful for practitioners to focus on the concept of validity in their informal assessment of clients. Counselors who purposefully analyze an accumulation of evidence will probably make more valid assessments of their clients.

L08 Item Analysis

This chapter includes information on both validity and item analysis. **Item analysis** focuses on examining and evaluating each item within an assessment. Whereas validity evidence concerns the entire instrument, item analysis examines the qualities of each item. Item analysis is often used when instruments are being developed or revised because it provides information that can be used to revise or edit problematic items or eliminate faulty items. Some of the simpler methods of item analysis can easily be used to evaluate items in classroom tests, questionnaires, or other informal instruments. There are also some extremely complex methods for analyzing items and this discussion will include a brief overview of the typical methods used to analyze items.

Item Difficulty

When examining ability or achievement tests, counselors are often interested in the difficulty level of the test items. The difficulty level of an item should indicate whether an item is too easy or, possibly, too difficult. **Item difficulty**, which is an index that reflects the proportion

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

of people getting an item correct, is calculated by dividing the number of individuals who answered the question correctly by the total number of people who answered the question:

$$p = \frac{\textit{number who answered correctly}}{\textit{total number of individuals}}$$

Item difficulty is often called *p* for short because it reflects the proportion or percentage of people who get an item correct (Urbina, 2014). An item difficulty index can range from .00 (meaning no one got the item correct) to 1.00 (meaning everyone got the item correct). Item difficulty does not really indicate difficulty; rather, because it provides the proportion of individuals who got the item correct, it shows how easy the item is. Let's use an example of an item on a counseling assessment test where 15 of the students in a class of 25 got the first item on the test correct. In this case, we would say the item difficulty (*p*) is .60:

$$p = \frac{15}{25} = .60$$

There are no clear guidelines for evaluating an item difficulty index, and the desired index difficulty level depends on the context of the assessment. In the previous example, we would need more information to interpret the index of .60 (interpretation of this item difficulty index might also depend on whether you are in the group who got the item correct or the group who got it wrong!). If the intent is to have items that differentiate among individuals, then a difficulty level of around .50 is considered optimum. Some assessment experts have suggested that in multiple-choice items, the probability of guessing needs to be considered in determining the optimum difficulty level (Cohen, Swerdlik, & Sturman, 2013). In other contexts, however, an instructor might be pleased with an item difficulty level of 1.00, which would indicate that everyone in the class was able to comprehend the material and answer the item correctly. Therefore, the determination of the desired item difficulty level depends on the purpose of the assessment, the group taking the instrument, and the format of the item. Evaluating an item difficulty index can frequently provide useful information to an instructor. For example, an easy item might have a low index because students find it confusing, whereas a difficult item may have a comparatively high item difficulty because the item may be written in a manner that students can figure out the correct response without knowing the content.

Item Discrimination

Item discrimination, or item discriminability, provides an indication of the degree to which an item correctly differentiates among the examinees on the behavior domain of interest. For example, have you ever taken a test and felt as if an item failed to discriminate between those individuals who studied hard and knew the material and those individuals who did not? Not only is item discrimination applicable to achievement or ability tests but it can also indicate if items discriminate in other assessment areas, such as personality and interest inventories. For instance, we may want to know if individuals who have been diagnosed as being depressed answered an item one way and if individuals who do not report depressive symptoms answer the same question another way. There are several methods for evaluating an item's ability to discriminate.

One of the most common methods for calculating an item discrimination index is called the *extreme group method.* First, examinees are divided into two groups based on high and low scores on the instrument (the instrument could be a class test, a depression inventory, or an assertiveness scale). The item discrimination index is then calculated by

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

subtracting the proportion of examinees in the lower group (lower %) from the proportion of examinees in the upper group (upper %) who got the item correct or who endorsed the item in the expected manner:

$$d = \text{upper \%} - \text{lower \%}$$

Item discrimination indices that are calculated with this method can range from +1.00 (all of the upper group got it right and none of the lower group got it right) to −1.00 (none of the upper group got it right and all of the lower group got it right). An item discrimination index with a negative number is a nightmare for test developers (particularly for instructors with vengeful students in the upper group who find out the item did not discriminate between those who studied and those who did not). Like item difficulty, interpretation of an item discrimination index depends on the instrument, the purpose it was used for, and the group taking the instrument.

The determination of the upper group and the lower group with item discrimination depends on the distribution of scores. Kelley (1939) showed that if there is a normal distribution, then optimal dispersion is accomplished by using the upper 27% for the upper group and the lower 27% for the lower group. Anastasi and Urbina (1997) suggested that with small groups, such as the typical classroom of around 30 students, the exact percentages to select each group need not be precisely defined but should be in the general range of 25–33%.

Another method for determining item discrimination is through the *correlational method*. These correlational approaches report the relationship between the performance on the item and a criterion. Often the criterion is performance on the overall test. The point biserial method and the phi coefficient are two commonly used correlational techniques. A thorough explanation of these techniques would be too lengthy for this discussion, but interested readers are directed to Crocker and Algina (1986). The result of these methods is a correlation coefficient that ranges from +1.00 to −1.00, with the positive and larger coefficients reflecting items that are better discriminators.

LO9 Item Response Theory

Before moving on to other topics, it is important to briefly address item response theory (IRT). Item response theory is a departure from classical test theory, which has been the dominant theory in assessment. Embretson and Hershberger (1999) argued that IRT is one of the most important recent developments in psychological assessment and that practitioners need knowledge about these approaches to assessment. **Item response theory**, or **latent trait theory**, builds on the concepts of item analysis. The basis for item response theory is different from classical test theory, where the focus is on arriving at one true score and our attempts to control error in the instrument. In classical test theory, the items on the instrument are a sample of the larger set of items that represents the domain and we get an indication of the individual's performance based on the combined items. In IRT, the focus is on each *item* and creating items that measure a particular ability or the respondent's level of a latent trait. IRT rests on the assumption that the performance of an examinee on a test item can be predicted by a set of factors called traits, latent traits, or abilities. Using an IRT approach, we get an indication of an individual's performance based not on the total score but, rather, on the *precise items* the person answers. As an example, with a traditional approach, we might say a score above 70 indicates that the student is reading at the sixth-grade level. With an item response approach, however, the student would need to answer certain items correctly to achieve a score indicating a sixth-grade reading level. The goal with this approach is to develop items that assess sixth-grade reading (not fifth- or seventh-grade reading). IRT involves

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the scaling of items, a procedure that requires quite sophisticated mathematical analyses. It should be noted that item response theory is not one approach to instrument development, but a family of approaches that focus on each item (Urbina, 2014). A thorough discussion of IRT is not appropriate for this book; interested readers are directed to Baker (2001), Van der Linden and Hambleton (1997), and Wilson (2005).

In IRT, there are two fundamental assumptions: *unidimensionality* and *local independence* (Henard, 2000). Unidimensionality means that the theoretical view is that each item will be constructed so that it measures only one ability or latent trait. This is difficult to meet in reality because of factors such as test anxiety or reading ability might influence an answer on personality assessment. Sometimes developers of assessments based on item response theory will use the term *dominant* factor, which acknowledges other factors may be slightly involved. The second assumption, local independence, requires an examinee's response on each item to be unrelated to responses on other items. Because the focus is on each item, it is assumed that the answer on one item does not influence an answer on another item. In IRT, the only factor that would influence item response is the latent trait (Henard, 2000).

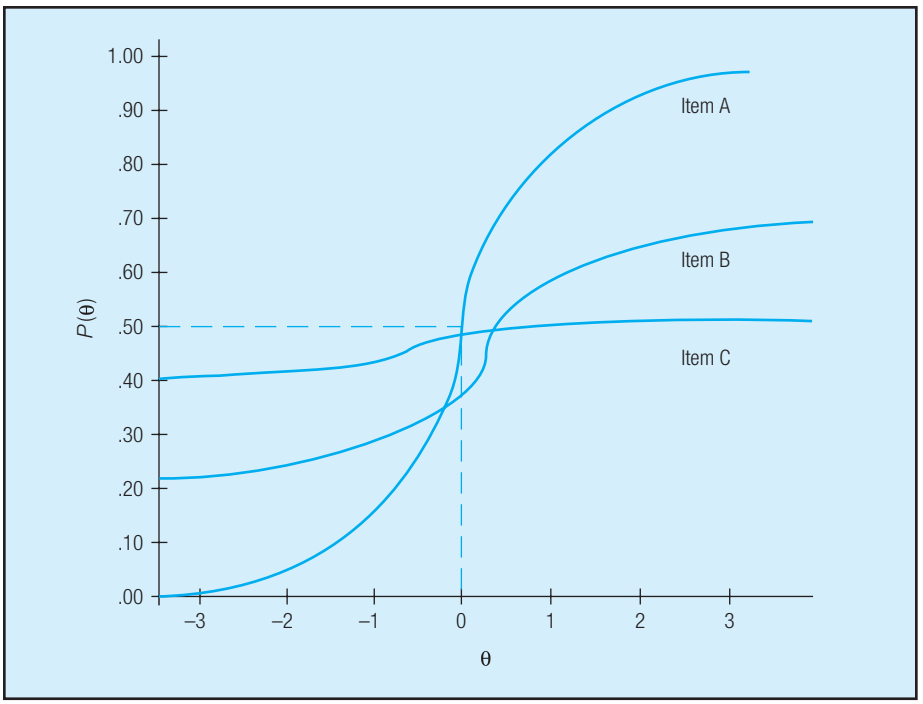
Item response theory suggests that the relationship between examinees' item performance and the underlying trait being measured can be described by a monotonically increasing function called an *item characteristic function* or an *item characteristic curve* (Baker, 2001). Although item response is used in assessment, such as personality inventories, where there is not a right or a wrong answer, it is most commonly used in assessment, such as achievement testing, where there is a right and a wrong answer. For the sake of simplicity, this example will assume that an ability is being assessed in which there is a right and a wrong answer. An assumption of item response theory is that each person taking the assessment has some amount of the underlying ability and the ability is often denoted by the Greek letter theta, θ (e.g., some level of reading ability). In item response theory, the ability is depicted on the abscissa as ranging from -3 to $+3$ with the average amount of the ability being indicated by a score of 0 (see Figure 4.4). Hypothetically, the scores range from $-\infty$ to $+\infty$, but rarely do the scores escape the bounds of -3 to $+3$ (Gregory, 2011).

To construct an item characteristic curve, an instrument developer starts by constructing an item that he believes will measure the ability of interest. The developer then gives the item to an appropriate group and then plots the results for each item using an item characteristic curve. As each person has a certain degree of the ability, the plotting is based on what is the probability that an examinee with that ability will get item correct. This probability will be denoted as $P(\theta)$. What is desired are items in which the probability of getting it correct is small for those individuals who have low abilities and the probability is large for those individuals with high abilities. Item A in Figure 4.4 has the item characteristic curve that is often desired as the probability of getting the item correct is low for those with little ability (i.e., -3, -2), but as the amount of ability increases, so does the probability of getting the item correct. In Figure 4.4, Item B is not a very good item because the probability of getting the item correct is approximately 20 percent with low levels of the ability and approximately 60 to 70 percent chance with high levels of the ability. With IRT, Item C is a particularly poor item as there is very little change or discrimination between those with low levels of the ability and those with high levels of the ability. Test developers using IRT want items that have a nice S shape on the item characteristic curve. In some instances, they want exactly what Item A represents in which those of average ability (0 on the horizontal axis) have a 50 percent probability of getting the item correct (.50 on the vertical axis).

For each item on an instrument, an item characteristic curve is calculated and there are two technical properties that describe the curve. The first is difficulty of the item as these difficulty levels correspond to the ability scale. The second technical property is the item discrimination, which concerns how well the item differentiates between those with

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 4.4
Hypothetical item characteristic curves



low ability and those with high ability. The slope of the curve provides an indication of discrimination as steep curves indicate that those with low ability got it wrong and those with high ability got it right (see Item A). A flat curve, such as Item C, indicates that there is little difference in terms of the probability of getting it right between those with low and high ability.

For the mathematically minded, there are three models that provide mathematical equations for item characteristic curves. Often instrument developers use the *two-parameter model,* which includes both difficulty and discrimination. The equation is:

$$P(\theta) = \frac{1}{1 + e^{-a(\theta - b)}}$$

Remember that $P(\theta)$ is probability that an examinee with a certain ability level (e.g., +1) will get item correct. The symbol e is the natural logarithm, which has a constant value of 2.71828. The mathematics of calculating a and b are quite complicated and beyond the scope of this book; however, once the instrument developer has these parameters, then the number can be plugged in for each θ level (i.e., −3, −2, −1, 0, 1, 2, 3) and the item characteristic curve can be plotted. If all the items in the instrument discriminate equally well, then the instrument developer can use the *one-parameter* or *Rasch* model, where a is not needed as the items discriminate equally:

$$P(\theta) = \frac{1}{1 + e^{-1(\theta - b)}}$$

The third model or the *three-parameter model* incorporates a parameter related to the probability of guessing. Interested readers are directed to Baker (2001). Once the item characteristic curve is plotted using one of the formulas, the instrument developers evaluate the curve to determine whether the item should be used. IRT can be very helpful in

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

developing a shorter version of an instrument that may have been developed using classical test theory because each item is evaluated regarding its ability to discriminate the level of the latent ability or trait.

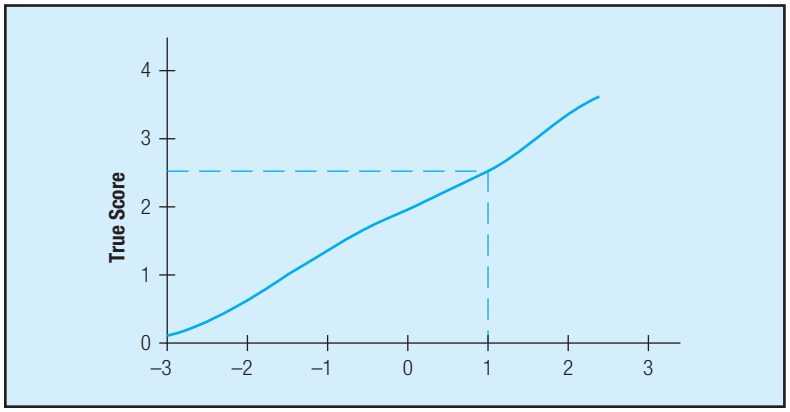
An advantage of IRT is that item parameters are *invariant* because their estimations are not tied directly to a sample. Therefore, an examinee's true score is calculated very differently with item response theory:

$$TS_j = \sum_{i=1}^{N} P_i(\theta_j)$$

in which TS_j is the true score for examinees with ability level θ_j, i denotes an item, and $P_i(\theta_j)$ is the probability estimate for that item at the desired ability level. Thus, as an example, let us say there is a 4-item test in which the probabilities for ability level 1 are .66 for item 1, .59 for item 2, .60 for item 3, and .71 for item 4. We would add those together to find the true score for someone at ability level 1. We would then add these together (.66 + .59 + .60 + .71 = 2.56), which would give us 2.56 as the true score for ability level 1. A *test characteristic curve* can then be calculated for each point on the ability scale, which means a true score can be computed for any point along the ability scale from negative to positive (Baker, 2001). Figure 4.5 is an example of a test characteristic curve that demonstrates that the true score for ability level 1 is 2.56. In the future, we can then use this test and know that those who score 2.56 have level 1 in terms of their ability level. Another advantage of IRT is that the reliability can be determined at different levels of the test characteristic curve.

There are numerous approaches to using item response theory in instrument development, and the methods continue to become more sophisticated (Embretson, 2010). Item response theory is influencing large-scale testing programs, particularly in the cognitive area. Another application concerns building item banks that include items scaled to different levels. These sophisticated item banks can provide useful information in terms of identifying individuals' academic strengths and limitations. Because of the extensive analysis and calibration of the items, IRT can be particularly useful in the development of criterion-referenced tests. It is also having an effect on **computer adaptive testing** in which performance on certain items influences the future items presented to the examinee. For example, if students cannot multiply a one-digit number by a one-digit number, it makes little sense to give them numerous questions about multiplying two-digit numbers by two-digit numbers. Thus, the computer sorts the questions based on previous responses. IRT can be particularly useful in achievement and ability adaptive testing, because the calibrated items can be used to quickly pinpoint an individual's ability level and identify possible gaps in knowledge. Item response theory is also serving as the theoretical base for

FIGURE 4.5
Test characteristic curve



Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

measures of personality and psychopathology. The approach is the same as achievement. We can find an example of using the IRT approach to analyze an instrument in Meijer, de Vries, and van Bruggen (2011), who examined the Brief Symptom Inventory-18. Another strength of an IRT approach is that it enables researchers to conduct rigorous tests of measurement across experimental groups (e.g., distressed versus nondistressed individuals). IRT approaches also have cross-cultural applications, where researchers compare the item characteristic curves of various ethnic/racial groups.

Summary

Validity concerns whether an assessment measures what it is designed to measure. An instrument is not validated; rather, it is the uses of the instrument that are validated. Validation evidence informs counselors on what they can infer from the results. A counselor is responsible for examining the accumulation of validation evidence and for determining how the instrument can be used and what can be inferred from the results. The field of measurement is moving away from the three traditional types of validity to an overall, integrated view of validation evidence. Therefore, the sources of evidence should not be viewed as separate and distinct types but, rather, as ways of accumulating evidence that contributes to overall evaluation of the appropriate uses and interpretations of the assessment. Many different techniques can be used in the process of gathering validation evidence. Counselors evaluating the validity of an instrument should examine what the preponderance of evidence indicates about whether that instrument can be used in a specific manner.

Items in well-constructed instruments are also analyzed. Two of the most common methods of item analysis involve determining the item difficulty and the item discrimination indices. Item response theory, which involves developing items that are calibrated to measure abilities or traits, is also having a substantial influence on instrument development.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 5

Ethical and Legal Issues in Assessment

Learning Objectives

- L01 Name resources for ethical decisions in assessment
- L02 Understand the portion of the ACA *Code of Ethics* on assessment
- L03 Explain clients' rights and responsibilities associated with assessment
- L04 Describe counselors' rights and responsibilities associated with assessment
- L05 Explain the Civil Rights Act of 1991 influence on assessment
- L06 Describe how the Americans with Disabilities Act Amendment Acts of 2008 affects assessment
- L07 Understand the Individuals with Disabilities Education Act of 2004 influence on assessment
- L08 Clarify how the Family Educational Rights and Privacy Act of 1974 affects assessment
- L09 Depict how the Health Insurance Portability and Accountability Act of 1996 influences assessment
- L010 Describe court cases that have influenced assessment practices

In considering the topic of assessment in counseling, it is important for counselors to consider ethical and legal practices. Knowledge of ethical and legal issues will help counselors from misusing instruments and possibly harming clients. Many times, misuses of assessments do not result from malicious behaviors; rather, these problems arise from ill-informed practitioners assessing and interpreting results in a careless manner. From the beginning of the assessment process, counselors need to understand both ethical and legal guidelines to practice in a professional and competent manner. Before discussing ethical and legal issues, it is important to ensure that counselors understand the differences between these two entities. Laws are related to a body of *rules*, whereas ethics are a body of *principles* that address proper conduct. Therefore, ethics and laws sometimes focus on different aspects or issues in assessment. The intent of both, however, is to protect people from the possible harm caused by the misuse of assessments. We will start with a discussion of ethics and then move to legal issues.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L01 Ethics

As we get started in our discussion of ethics, it is important to point out a few ethical resources that may be helpful to clinicians. Due to the complexities of assessment issues in counseling, a number of sources are available to aid practitioners in practicing in an ethical manner. Table 5.1 includes a listing of some of the primary resources related to making ethical decisions in assessment. The ethical standards of professional organizations related to counseling and assessment are the foundational source for these materials. In the American Counseling Association's (2014) *Code of Ethics*, an entire section is devoted to evaluation, assessment, and interpretation (this section is reproduced in Appendix A of this book) and this section of the ethical code is summarized in the next portion of this chapter. There are also ethical guidelines specific to school counseling (American School Counselor Association, 2010) and career counseling (National Career Development Association, 2007). Other professional organizations have standards that address assessment issues, such as the American Psychological Association's (2010) *Ethical Principles of Psychologists and Code of Conduct*. These ethical standards and codes establish principles and define ethical conduct for members of the organizations. Members of these organizations are required to adhere to their ethical standards, which have arisen out of clients being harmed or the potential for clients to be harmed.

Another major resource regarding both ethical and legal questions is the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), which is often referred to as the *Standards*. The *Standards* serves as a technical guide for testing practices and concerns the appropriate uses of tests. The *Standards* provides expectations of professional services and should be reviewed by every practitioner who is involved in any type of assessment. The *Standards* has also been used in legal cases as an indicator of standards of practice related to assessment in counseling. Two sections of the *Standards* are summarized in this chapter. The first section will discuss client rights and responsibilities with a focus on client right. The second will discuss counselor rights and responsibilities with a focus on counselor responsibilities.

Another source related to ethical practices in assessment is the *Standards for the Qualification of Test Users* (American Counseling Association, 2003), which synthesizes the professional qualifications essential when using tests in counseling.

Another excellent resource related to ethical issues in assessment is the *Responsibilities of Users of Standardized Tests* (Association for Assessment in Counseling, 2003), which is often referred to as the *RUST Statement*. The *RUST Statement* is intended to address the

TABLE 5.1
Resources for ethical practices in counseling assessment

- American Counseling Association (2014). *ACA code of ethics.* Alexandria, VA: Author.
- American Counseling Association (2003). *Standards for qualifications of test users.* Alexandria, VA: Author.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.
- American Psychological Association (2010). *Ethical principles of psychologists and code of conduct.* Washington, DC: Author.
- Association for Assessment in Counseling (2003). *Responsibilities of users of standardized tests (RUST)* (3rd ed.) Alexandria, VA: Author.
- Association for Assessment in Counseling and Education. (2012). *Standards for multicultural assessment: Fourth revision.* Alexandria, VA: Author.
- Eyde, L. D., Robertson, G. J., & Krug, S. E. (2010). *Responsible test use: Case studies for assessing human behavior* (2nd ed.). Washington, DC: American Psychological Association.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

needs of members of the American Counseling Association and to promote accurate, fair, and responsible use of standardized tests. The focus of this document is not on the development of instruments but rather on the proper use of tests and methods for safeguarding against the misuse of tests. The *RUST Statement* is currently being rewritten so readers should look for a new edition of the statement soon.

A fourth resource that practitioners find extremely helpful is the casebook entitled *Responsible Test Use: Case Studies for Assessing Human Behavior-Second Edition* (Eyde, Robertson, & Krug, 2010). This resource includes 85 cases illustrating both proper and improper test usages. Examples relate to such topics as professional development, test selection, test administration, test interpretation, and reporting test results to clients and administrative policy issues. This resource is particularly useful in describing real-world situations and applying ethical principles to these actual incidents. The casebook's major emphasis is on the application of basic principles of sound test selection, use, and interpretation. It contains 64 elements of good test use to help readers learn about proper test usage through the case studies.

There are a number of other helpful resources for counselors involved in assessment. Also of primary importance is the Association for Assessment and Research in Counseling (http://aarc-counseling.org/), which is affiliated with the American Counseling Association. The Association for Assessment and Research in Counseling, previously named the Association for Assessment in Counseling and Education, has developed a number of standards related to assessment and certain counseling specialties. For example, there are the *Standards for Assessment in Career Counseling, Standards for Assessment in Marriage, Couple, and Family Counseling, Standards for Assessment in Mental Health Counseling, and Standards for Assessment in Substance Abuse Counseling*. Interested readers are directed to the webpage of the Association for Assessment and Research in Counseling for these important standards.

L02 ACA Code of Ethics

As mentioned above, the *ACA Code of Ethics* (ACA, 2014) is an important resource for counselors involved in assessment. In fact, the importance of assessment in the counseling field is documented by an entire section (i.e., Section E) of the ethical code being devoted to evaluation, assessment, and interpretation. This section of the *ACA Code of Ethics* (ACA, 2014) can be found in Appendix A and the following is a brief overview of Section E of the *ACA Code of Ethics* (ACA, 2014).

Introduction

Consistent with this book, Section E of *ACA Code of Ethics* (ACA, 2014) begins with the premise that "Counselors use assessment as one component of the counseling process, taking into account the clients' personal and cultural context" (p. 11). This introductory section goes on to instruct counselors to promote the well-being of clients or groups of clients by creating and using appropriate assessments. Specifically these ethical codes call for the appropriate use of "educational, mental health, psychological, and career assessments" (p. 11).

E.1. General. In this section of the ethical code, the authors start with the basic purpose of assessment, which is stated as "The primary purpose of educational, mental health, psychological, and career assessment is to gather information regarding the client for a variety of purposes, including, but not limited to, client decision making, treatment planning, and forensic proceedings" (p. 11). The ethical code further stipulates that assessments can involve either qualitative or quantitative methodologies. The second portion in this general section addresses the important topic of client welfare and clearly articulates that counselors do not misuse assessment results or interpretation. In addition, the ethical code

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

requires counselors to take reasonable steps to prevent others from misusing the information provided by assessments. Furthermore, counselors need to respect the rights of clients to receive assessment results and interpretation and, further, emphasizes that clients have the right to understand the bases for counselors' recommendations and conclusions.

E.2. Competence to Use and Interpret Assessment Instruments. The ethical code clearly states that counselors must know the limits of their competencies in the area of assessment and that counselors only use the testing and assessment services for which they have been adequately trained and are competent to use. Although test publishers often have user qualifications that are designed to restrict unqualified users from purchasing instruments for which the counselor is not trained, the ultimate responsibility for competence use rests with the counselor. Often instruments are categorized as A, B, or C level assessments and the following is an example of this classification:

QUALIFICATION LEVEL A:

There are no special qualifications to purchase these products.

QUALIFICATION LEVEL B:

Tests may be purchased by individuals with:

A master's degree in psychology, education, occupational therapy, social work, or in a field closely related to the intended use of the assessment, and formal training in the ethical administration, scoring, and interpretation of clinical assessments.

OR

Certification by or full active membership in a professional organization (such as ASHA, AOTA, AERA, ACA, AMA, CEC, AEA, AAA, EAA, NAEYC, NBCC) that requires training and experience in the relevant area of assessment.

OR

A degree or license to practice in the healthcare or allied healthcare field.

OR

- Formal, supervised mental health, speech/language, and/or educational training specific to assessing children, or in infant and child development, and formal training in the ethical administration, scoring, and interpretation of clinical assessments.

QUALIFICATION LEVEL C:

Tests with a C qualification require a high level of expertise in test interpretation, and can be purchased by individuals with:

A doctorate degree in psychology, education, or closely related field with formal training in the ethical administration, scoring, and interpretation of clinical assessments related to the intended use of the assessment.

OR

Licensure or certification to practice in your state in a field related to the purchase.

OR

- Certification by or full active membership in a professional organization (such as APA, NASP, NAN, INS) that requires training and experience in the relevant area of assessment.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

This book is intended for students in master's-level counseling programs and it is designed to assist such students in learning about the area of assessment and to prepare to administer and interpret the results of Level B assessments. In the *ACA Code of Ethics* (ACA, 2014), the section on competence use and interpretation further explores the topic of counselors use of technology-assisted interpretation. As we will see in future chapters of this book, some assessment instruments now have interpretive reports that are produced by computers. A counselor cannot ethically use these technology-assisted interpretations unless they have specific training in the construct being measured (e.g., vocational interests). Furthermore, counselors must also take reasonable steps to ensure that any one the counselor is supervising also uses assessment techniques appropriately. Moreover, counselors are responsible for the appropriate "application, scoring, interpretation, and use" (ACA, 2014, p. 11) of instrument no matter if they do the scoring and interpretation themselves or if they use technology or other services to provide this information to clients. Counselors must also have a thorough understanding of the psychometrics of the instruments used when either decisions are made or policies are invoked as a results of assessments.

E.3. Informed Consent in Assessment. Informed consent is particularly important in assessment as client may not readily understand which constructs are being measured or how the assessment results will be used. Prior to taking an assessment, the *ACA Code of Ethics* (ACA, 2014) clearly state that counselors explain to clients the nature and purposes of the assessment and the specific use of results. This explanation should also be given to the client or other legally authorized personnel in terms and language that the individual can understand. For minors, typically the parents or guardian will need to be the one who authorizes or approves of the assessment.

In providing assessment results, the counselor must consider the client's or examinee's welfare, explicit understandings, and any other prior agreements regarding who specifically can receive the assessment results. In general, counselors cannot divulge any assessment results to a third party without written permission from the client or the client's legal representative (e.g., parents). Counselors must also ensure that only accurate and appropriate interpretations are released regarding any individual or group.

E.4. Release of Data to Qualified Personnel. Sometimes counselors will release assessment data to other qualified personnel, such as another clinician the client will be seeing or in the case of a school counselor to a university admissions officer. A counselor can only release assessment data in which a client's identity is revealed with the consent of the client or the client's legal representative. It is usually required that this consent be in writing and that the assessment data be released only to persons recognized by the counselor as being able and qualified to interpret the assessment information.

E.5. Diagnosis of Mental Disorders. As will be explored in other chapters of this book, diagnosis of a mental disorder is often a part of the assessment process. The *ACA Code of Ethics* (ACA, 2014) requires that counselors take special care in properly diagnosing clients with mental disorders. The ethical standards further stipulate that counselors carefully select and appropriately use any assessment technique (including personal interviews) to determine client care, which includes focus of treatment, type of treatment, and types of follow-ups. Furthermore, counselors are required to recognize that culture influences "the manner in which clients' problems are defined and experienced" (ACA, 2014, p. 11). Therefore, counselors should consider cultural and socioeconomic factors when diagnosing clients with mental disorders. Counselors should also recognize historical and social prejudices that have resulted in the misdiagnosis and pathologizing of certain individuals or groups. Counselors should also be aware of such biases and strive to address these biases in themselves or others. Of particular importance, counselors "may refrain from making and/or reporting a diagnosis if they believe that it would cause harm to the client or others"

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(ACA, 2014, p. 11). In general, counselors are instructed by the ethical standards to carefully consider both the positive and negative implications involved in any diagnosis.

E.6. Instrument Selection. As the previous chapters in this book have emphasized, considering an instrument's psychometric qualities is vital to selecting an instrument. The *ACA Code of Ethics* (ACA, 2014) agree with this approach and direct the counselor to carefully consider an instrument's reliability, validation evidence, psychometric limitation, and appropriate use when selecting assessment instruments. The ethical standards also encourage the counselor to, whenever possible, use multiple assessment sources and/or instruments when the counselor is "forming conclusions, diagnoses, or recommendations" (ACA, 2014, p. 11). I concur with this standard and strongly recommend that counselors use multiple sources of assessment information in their professional work with clients. A counselor may refer a client to another professional or a third party (e.g., a psychologist) for assessment. In these referral cases, the counselor should provide detailed information about the referral question(s) and sufficient client information in order to facilitate the selection and use of appropriate assessment instrument.

E.7. Conditions of Assessment Administration. Major problems can occur if an instrument is administered inappropriately; therefore, the *ACA Code of Ethics* (ACA, 2014) require that counselors "administer assessments under the same conditions that were established in their standardization" (p. 11). As will be noted when we discuss legal issues, sometimes accommodation in the instrument's administration will need to be made for individuals with disabilities and these accommodations will need to be noted. Also, sometimes there will be unusual behaviors or irregularities during the administration of an assessment and these conditions should also be noted in the interpretation. Occasionally these accommodations or irregularities may result in the assessment results being invalid or of questionable validity.

Increasingly, instruments are being administered by computers and counselors have the responsibility of ensuring that "technologically administered assessments function properly and provide clients with accurate results" (ACA, 2014, p. 11). If the instrument was not originally developed to be administered by a computer, the counselor is charged with investigating whether the technology-assisted administration produces equivalent results to the original assessment. Also, counselors do not permit clients to take assessment instruments unsupervised unless that assessment instrument was designed, intended, and validated for self-administration and/or scoring.

E.8. Multicultural Issues/Diversity in Assessment. As demographics are changing and counselors are increasingly working with diverse clients, it is particularly important that they attend to issues of multiculturalism in their assessment practices. If the instrument is norm-referenced, counselors must closely examine the information provided on the norming group as the *ACA Code of Ethics* (ACA, 2014) recommend that "counselors select and use with caution assessment techniques normed on populations other than that of the client" (p. 12). Multiculturalism is a broad term and the *ACA Code of Ethics* (ACA, 2014) call on counselors to consider assessment issues related to age, color, culture, disability, ethnic group, gender, race, language preference, religion, spirituality, sexual orientation, and socioeconomic status. Specifically, the ethical standards recommend that counselors recognize the effects of these factors on the administration and interpretation of the instrument, and that counselors interpret test results in the proper perspective with other relevant factors.

E.9. Scoring and Interpretation of Assessments. It is of paramount importance that the counselor consider the client's personal and cultural background when reporting assessment results. In the reporting of assessment results, counselors should also consider the level of the client's understanding of the results and the impact of the results on the client. When appropriate, the counselor also notes reservations about the results based on reliability information, validation evidence, or limitations of the norming group.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Therefore, it is critical that the counselor be aware of issues related to reliability and validity. Furthermore, counselors should be cautious in their interpretation of results when assessment instruments have limited psychometric information or insufficient empirical data. When there is limited empirical data, the counselor should explicitly communicate with the examinee the specific usage(s) of the instrument. When using instruments or assessments with questionable reliability or validity, the counselor should qualify any conclusions, diagnoses, or recommendations that are based on such instruments. Counselors who provide assessment services in terms of scoring and interpreting results confirm the validity of these interpretations. Furthermore, in providing assessment services, counselors "accurately describe the purpose, norms, validity, reliability, and applications of the procedures and any special qualifications applicable to their use" (ACA, 2014, p. 12). The ethical code makes it quite clear that at all times, counselors must maintain their ethical responsibilities to those being assessed.

E.10. Assessment Security. Not only do counselors have an ethical responsibility in regards to the administration, scoring, and interpretation of assessment results, but they also have the responsibility to "maintain the integrity and security of tests and assessments consistent with legal and contractual obligations" (ACA, 2014, p. 12). This means that counselors refrain from appropriating, reproducing, or modifying published instruments or parts of assessment instruments without the acknowledgement and permission from the publisher.

E.11. Obsolete Assessment and Outdated Results. It is not appropriate to use an outdated or obsolete version of an instrument (e.g., earlier edition of an instrument). As we will see, many commonly used instruments have been updated and counselors need to use the most current version of the assessment. Furthermore, counselors also have an ethical responsibility to ensure that others are not using outdated or obsolete versions of an assessment instrument.

E.12. Assessment Construction. Many times counselors are involved in the development of new assessment in field of counseling. In constructing new assessments, the *ACA Code of Ethics* (ACA, 2014) instructs counselors to use "established scientific procedures, relevant standards, and current professional knowledge for assessment design in the development, publication, and utilization of assessment techniques" (p. 12).

E.13. Forensic Evaluation: Evaluation for Legal Proceedings. According to the Merriam-Webster (2014) dictionary, *forensic* is defined as (a) relating to the use of scientific knowledge or methods in solving crimes; or (b) relating to, used in, or suitable to a court of law. According to the *ACA Code of Ethics* (ACA, 2014), "when providing forensic evaluations, the primary obligation of counselors is to produce objective findings that can be substantiated based on information and techniques appropriate to the evaluation, which may include examination of the individual and/or review of records" (p. 12). Counselors form professional opinions based on their specialized knowledge and expertise that can be supported by the detailed information gathered in the evaluation. Counselors need to define the limits of their testimony or reports, particularly when an examination of the individuals has not been conducted. Counselors need to provide in writing to individuals being evaluated that the professional relationship is evaluative and that it is not therapeutic. Furthermore, counselors should provide in writing who will receive the evaluative report. Informed consent can be murky in forensic evaluations as individuals may be mandated to participate in an evaluation. Nevertheless, counselors who perform forensic evaluations must obtain written consent from those being evaluated or from their legal representative unless a court orders an evaluation to be conducted without the written consent of the individuals being evaluated. When evaluating children or adults who may lack the capacity to give consent, counselors obtain written informed consent from the parents or legal

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

guardians. Counselors should not evaluate "current or former clients, clients' romantic partners, or clients' family members for forensic purposes" (ACA, 2014, p. 12). Moreover, counselors do not provide therapeutic or counseling services to individuals for whom they are evaluating for forensic purposes. In addition, according to the *ACA Code of Ethics* (ACA, 2014), counselors who provide forensic evaluations avoid potentially harmful professional or personal relationships. This includes not providing forensic evaluations of family members, romantic partners, and close friends of individuals they are evaluating or have evaluated previously.

L03 Client Rights and Responsibilities

In considering the ethical practices in assessment, the *ACA Code of Ethics* (ACA, 2014) provides a primary resource for counselors, but another prominent resource is the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), which includes a chapter devoted to the rights and responsibilities of test takers. It should be remembered that the *Standards for Educational and Psychological Testing* are often referred to as the *Standards* and they use the term *test* rather than *instrument*. I will use the terms *tests*, *instruments*, and *assessments* interchangeably here. There also may be some repetition of information covered above, but this repetition is intentional since ethical practice is so important.

In considering ethical practices, it is important to consider the rights of clients in the testing process. The *Standards* indicate that it is incumbent on counselors to explain to clients their rights and responsibilities associated with assessment. The *Standards* clearly indicate that clients have the right to be assessed with assessments that meet current professional standards. Furthermore, clients have the right to basic information about the instrument and how the results will be used. According to the *Standards* (AERA, APA, & NCME, 2014), "fair and equitable treatment of test takers involves providing information about the general nature of the test, the intended use of test scores, and the confidentiality of the results in advance of testing" (p. 131). Clients or their legal representative need sufficient information to make an informed decision regarding participation. Furthermore, the greater the consequences of assessment, the more fully that a client should be informed about the assessment and the more focus there is on voluntary consent except when assessments are legally mandated or required by the courts. If the assessment is optional, clients also have the right to know the consequences of not taking the assessment. Moreover, clients have the right to understand how results will be disseminated and "they also have the right to protection of their personally identifiable score results from unauthorized access, use, or disclosure" (p. 133). The dissemination of assessment information may be applicable to privacy laws, which will be discussed in a later section of this chapter.

The *Standards* (AERA, APA, & NCME, 2014) also stipulates that informed consent should be obtained from either the client or their legal representative before the assessment process begins except for certain circumstances. These circumstances are (a) when assessment without consent is mandated by law or governmental regulation, (b) when testing is performed as a part of regular school activities, or (c) when consent is clearly implied (e.g., employment testing). Furthermore, informed consent *involves* providing assessment information in a language clients can clear understand.

The *Standards* also address practice materials available to clients and state that practice materials should have the same appearance and format as the actual test or assessment (e.g., web-based assessment should have practice materials available online). Any information disseminated to a client should be released to all clients, and in the cases of clients with disabilities, it may be required by law that this information is disseminated in a format that is accessible. Information about assessment policies (e.g., accommodation policies) also should be made available to all examinees. It should be noted, however, that requests for

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

accommodation are the responsibility of the client or in the case of minors, the client's guardian. In addition, clients have the right to know policies related to retesting if there is a major disturbance during the administration, if the client believes the results are not reflective of his or her capabilities, or if the client wants to take the assessment again after improving his or her knowledge, abilities, or other personal characteristics.

In addition to rights regarding assessment, clients also have responsibilities. According to the *Standards* (AERA, APA, & NCME, 2014), the responsibilities include "being prepared to take the test, following the directions of the test administrator, representing themselves honestly on the test, and protecting the security of the test" (p. 132). This section addresses how cheating can influence the validity of the interpretation and how divulging any information about the instrument is problematic. Furthermore, the *Standards* state that if testing programs use specific procedures to detect irregularities in testing (e.g., examining response speed, comparing present results to previous results), then examinees must be made aware of such possibilities before taking the assessment.

L04 Counselors Rights and Responsibilities

Many of the counselor's responsibilities regarding assessment are related to ensuring that clients are informed of their rights and responsibilities; however, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) also include a chapter that specifically addresses the test user rights and responsibilities. The general standard for this section of the *Standards* (AERA, APA, & NCME, 2014) is as follows:

> *Test users are responsible for knowing the validity evidence in support of the intended interpretations of scores on tests that they use, from test selection through the use of scores, as well as common positive and negative consequences of test use. Test users also have a legal and ethical responsibility to protect the security of test content and the privacy of test takers and should provide pertinent and timely information to test takers and other test users with whom they share test scores. (p. 142)*

The standards contained in the chapter related to test users or counselors responsibilities are presented in three thematic clusters that address (a) validity of interpretation, (b) disseminating information, and (c) instrument security and protection of copyrights.

In considering validity of interpretation, the counselor must first consider their qualifications and what qualifications are listed in the manual related to the administration and interpretation of the instrument. Ultimately, the counselor is responsible for ensuring that the clinical training requirements, ethical codes, and legal standards for administering and interpreting assessments are met. Some states have laws related to which instruments counselors can use and counselors are responsible for knowing these laws. Also related to validity of interpretation, the *Standards* state that it is the counselor's responsibility to be knowledgeable of the instrument's manual and appropriate research so that the counselor is aware of both appropriate and inappropriate usages and for whom the instrument is appropriate or inappropriate. Furthermore, counselors should be clear about the reasons for why they are using an instrument and should periodically re-verify that they are using the instrument appropriately and with suitable clients. Counselors should also be aware of potential scoring errors and potential misuses of assessments and take appropriate actions related to these problems. Counselors also do not interpret instrument results in isolation and, when possible, corroborate assessment results with other information. Moreover, counselors have the responsibility for informing appropriate clients that accommodation may be available and for ensuring that accommodations are appropriate. In fact, the *Standards* state that counselors should be able to cite the information or evidence (e.g., instrument manual or research study) used to select the appropriate accommodations (AERA, APA, & NCME, 2014).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In considering their responsibilities related to assessment, counselors also consider the dissemination of information. As has been mentioned numerous times, counselors inform others of appropriate assessment uses, how the assessment will be administered, factors associated with scoring, how the scores will be used, how long the records will be maintained, and to whom and under what circumstances the records may be released (AERA, APA, & NCME, 2014). Unless there are legitimate circumstances that require the withholding of results, a counselor is obligated to provide a timely interpretation of the results to the client or to others entitled to the results (e.g., parents). When results are disseminated to a client, the public, or policy makers, the counselor should provide supplemental information to minimize possible misinterpretation (e.g., the difference between percentage correct and percentile scores). If a client or a counselor are concerned about the integrity of the results, then the counselor must inform the client of his or her rights, which could include the appeal process or representation by counsel. Counselors should inform clients if there is a possibility of retaking the assessment and the procedures and policies involved in that process. Of paramount importance is the protection of clients' and, in some cases, institutions' privacy; thus, counselors do not disseminate assessment information unless it is agreed upon or authorized by law.

In the area of counselor rights and responsibilities, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) also address the topics of test security and protection of copyrights. As the validity of interpretation is influenced by instrument security, counselors must ensure that assessment content, scoring, and information necessary for interpretation are kept secure. Keeping information secure also applies to previous editions of instruments. Furthermore, counselors may not reproduce or create electronic versions of copyrighted materials for routine use without consent from the copyright holder.

In conclusion, the ethical use of assessment interventions is a professional responsibility. Counselors are responsible not only for their own professional conduct but also for taking reasonable steps to prevent other professionals from misusing assessment procedures or results. Certainly, there have been misuses of assessment devices in the past; however, if professionals monitor their own and others' use of assessments, then future misuses should be significantly reduced. It should be remembered that ethical information is one source for advising counselors on best practices in assessment, but counselors should also be aware of legal issues that may influence their practices.

Legal Issues in Assessment

As mentioned earlier, principles and ethical standards direct professional behavior in assessment. There are also *laws* to which professionals need to adhere. Laws emanate from two sources: legislation and litigation. *Legislation* concerns governmental bodies passing laws. For example, many state legislatures have passed bills related to the licensing of Licensed Professional Counselors. *Litigation* is the other way that laws are formed by "rules of law." In this case, courts interpret the U.S. Constitution, federal law, state law, and common law in a particular case, and their ruling then influences the interpretation of the relevant laws. Both legislation and litigation have had an effect on assessment. In the past, most of the legislation and litigation that influenced assessment concerned employment testing and educational testing. This area, however, is always evolving and changing—a new law could be passed tomorrow or a court could make an influential decision next week. Therefore, counselors must continually strive to keep abreast of these evolving legal issues.

Legislation

The U.S. Congress has passed a number of influential legislative acts related to assessment. Most of these laws were not written for the specific purpose of controlling assessment practice, but they do contain content and language that are related to testing practice. Because these are federal laws, rather than state laws, these acts affect the entire country of the United States.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L05 ***Civil Rights Act of 1991.*** Title VII of the Civil Rights Acts of 1964 and the 1972 and 1978 amendments of this act outlawed discrimination in employment based on race, color, religion, sex, or national origin. The original legislation created the Equal Employment Opportunity Commission (EEOC), which was charged with developing guidelines to regulate equal employment. In the 1970s, the EEOC developed very strict guidelines related to the use of employment tests, addressing the type of reliability and validity evidence that needed to be present. During the 1980s, there was considerable controversy about the EEOC. This controversy is partly what led to the passing of the Civil Rights Act of 1991.

Several controversial court decisions that certain civil rights advocates found disturbing also may have influenced the passage of the Civil Rights Act of 1991. These decisions, during the 1980s, were changes that resulted from the landmark case *Griggs v. Duke Power Company* (1971). This case involved African American employees of a private power company. The employees filed suit against the power company, claiming that the criteria for promotion, such as requiring a high school diploma and passing scores on the Wonderlic Personnel Test and the Bennett Mechanical Comprehension Test, were discriminatory. The Supreme Court ruled in favor of the employees and found that although an instrument may not appear biased, it may discriminate if it has a disparate or adverse impact. The Court also ruled that if there is disparate impact, the employer must show that the hiring procedures are job-related and a reasonable measure of job performance. Because the validity of the two instruments was an issue in this decision, this case spurred a stronger focus on the validity of employment tests in making employment decisions.

One of the first cases that diverged somewhat from the *Griggs* case was *Watson v. Fort Worth Bank and Trust* (1988). In this case, the Supreme Court did not rule on behalf of the African American woman who had been passed over for a supervisory position. This ruling raised questions about employment selection procedures for supervisory positions. The findings indicated that, for supervisory positions, employment testing procedures do not need strong validity evidence and that subjectivity can enter into these decisions. In *Wards Cove Packing Company v. Antonio* (1989), the Supreme Court ruled on the concept of adverse or disparate impact that some employers thought the Court would rule on in the *Watson* case. In *Wards Cove,* most of the cannery jobs were held by Eskimos and Filipinos, whereas the company's noncannery jobs were held by European Americans. The judicial decision in this case shifted the burden to the plaintiff to show that there are problems with the selection procedures. Based on this case, plaintiffs would need to show that the instruments were not job related, which is difficult to do with limited access to human resources or personnel information and records.

In many ways, the Civil Rights Act of 1991 codified the principles of *Griggs v. Duke Power Company.* The plaintiff must show that there is an adverse or disparate impact. Then the burden shifts to the employer to show that discriminatory hiring procedures are needed because of business necessity and job relatedness (Hagen & Hagen, 1995). Therefore, the Civil Rights Act shifts more of the burden back to the employer and requires that hiring procedures and employment tests be connected to the duties of the job.

Another feature of the Civil Rights Act of 1991 is the ban on separate norms in employment tests. The act specifically states that it is unlawful to adjust the scores of, use different cutoff scores for, or otherwise alter the results of employment-related tests because of race, color, religion, sex, or national origin. Some individuals in the assessment field believe that this portion of the act may have untold ramifications. The focus of the ban was on separate norms used by the U.S. Employment Services (USES) for minority and nonminority job applicants. In the 1980s, the USES sent potential employees' General Aptitude Test Battery (GATB) percentile scores to employers for them to use in hiring decisions. A problem with this procedure was that African Americans and Hispanics tended to have lower average scores on the GATB, and, thus, their percentile scores were not competitive.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The USES decided to use separate norms so that African Americans would be compared with other African Americans. The public outcries, however, led to the ban of separate norms in 1991 (Sackett & Wilk, 1994). In addition to issues of race, there are other concerns about whether to use separate norms, such as for men and women on personality tests. Some findings indicate that women are more likely to disclose negative information about themselves, which may be problematic when personality instruments are used for hiring. The debate surrounding the advantages and disadvantages of separate norms will most likely continue in the coming years. Racial differences in hiring and performance on employment tests are complex and multidimensional issues. Gottfredson (1994) suggested that the approach to these issues needs to be broadened. She contended that employment tests did not cause the racial differences, and we cannot expect that adjusting or not adjusting the scores will solve these complex problems.

L06 ***Americans with Disabilities Act Amendment Acts of 2008.*** Similar to the Civil Rights Act of 1991 that restored some of the original intent of earlier versions of the act, the American with Disabilities Act Amendment Acts of 2008 sought to restore some of the original language of the Americans with Disabilities Act (ADA) of 1990. In particular, there had been some court cases that narrowed the definition of disabilities and the American with Disabilities Act Amendment Acts (ADAAA) sought to restore the original definition of disabilities, which is physical or mental impairment that substantially limits a major life activity; a record of such an impairment; or being regarded as having such an impairment. As a result of ADAAA, more Americans are likely to qualify as being disabled and to be further protected from discrimination (Scott, 2010). ADAAA bans discrimination in employment and in access to public services, transportation, and telecommunications on the basis of physical and mental disabilities. The precise differences between ADA and ADAAA in terms of the definition of disability are beyond the scope of this book. Interested readers are directed to Scott (2010) for a concise overview of those changes. ADA and ADAAA came to address testing through the focus on employment and employment testing, and determining whether someone has a record of being disabled. I will not be discussing testing related to documenting if someone has a disability as most counselors will not be involved in the assessment of individuals to determine if they have a mental disability; that assessment is usually conducted by a psychologist (Warren, 2011).

The language of ADA and ADAAA is broad and covers multiple testing situations. Specific to employment testing within ADAAA, employers are only allowed qualification standards, employment tests, or other selection criteria IF the employer can demonstrate that the requirement is job related to the position in question AND consistent with business necessity.

For school counselors, the ADAAA definition of disability also covers students who may qualify for Section 504 of the Rehabilitation Act of 1973. Section 504 prohibits discrimination on the basis of disability in programs and activities that receive federal funds and requires that reasonable accommodation for the testing of students with disabilities be made. A student who qualifies for Section 504 may not necessarily qualify for special education (see the next section, "Individuals with Disabilities Education Act of 2004") but will qualify for reasonable and appropriate accommodation in educational testing. This language raises some troublesome questions related to assessment. First, how are reasonable and appropriate accommodations defined? As of yet, there is still some ambiguity related to what constitutes reasonable and appropriate accommodation. In addition, there are thousands of disabilities, and, therefore, it is difficult for school counselors to make recommendations on precisely how to make accommodations that are reasonable for all possible situations. A counselor must be careful with any nonstandard administration of a standardized instrument because standardized instruments are designed to be administered in a uniform

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

manner. As indicated earlier in the discussion of ethics, the ultimate responsibility for the appropriate use of an instrument rests with the user, so clinicians have a responsibility to investigate what are appropriate accommodations and research studies and assessment websites are often valuable resources.

LO7 ***Individuals with Disabilities Education Act of 2004.*** Another piece of legislation affecting assessment and testing of individuals with disabilities is the Individuals with Disabilities Education Act (IDEA), which was last reauthorized in 2004. IDEA is a law ensuring educational services to children with disabilities. Many different aspects of IDEA are pertinent to counselors, particularly school counselors. However, the focus of this discussion will be on *assessment*. IDEA legislates that each state must have a comprehensive system that can identify, locate, and evaluate children ages birth through 21 who may have a disability. The request for evaluation can come from either the parent or a public agency (e.g., the child's school) when there is a suspicion that there may be some sort of disability. If a public agency suspects there is a disability, the agency must provide prior written notice to the parents explaining why they wish to conduct the evaluation, a description of each assessment, and a description of procedural safeguards (e.g., where parents can go to receive more information about the process). Once the parents have consented to the evaluation, the evaluation must be conducted within 60 days unless there is other legislation at the state level. A parent's request for an evaluation can be refused by the public agency if they believe there is not a disability and there is a system for parents to appeal this decision. The intent of the evaluation is not just to identify whether a child has a disability; rather, the evaluation is designed to *fully learn* the nature and extent of the special education and related services the child needs. If a child has a disability, then the public agency must develop an **Individualized Education Plan (IEP)**, which is developed by a team of educators and the parents. The evaluation must be conducted individually. Table 5.2 includes specifics about how the evaluation must be conducted.

Often these evaluations will be conducted by school psychologists, but school counselors are often on an IEP team, which will have input into the evaluation and subsequent decisions about the appropriate educational plan. In some cases, mental health counselors may be working with children or adolescents who have been assessed, and these professionals may ask the parents to have the results released to them. If you examine Table 5.2, you will notice that a significant portion of the federal regulations are consistent with best practices in assessment. For example, as has been emphasized in this book, the evaluation should not be based on one test but should include a variety of assessment tools. Typically, information is gathered using a variety of techniques (e.g., observation, interviews) and from a variety of sources (e.g., the student, parents, and teachers). The evaluation should include instruments that are technically sound in all areas in which there is a suspected disability. As also can be seen from Table 5.2, there is considerable focus on issues of race, culture, and language so that the assessments are not discriminatory. The assessments should be given in the child's native language or in another mode or form that is likely to yield accurate information on what the child knows and can do academically, developmentally, and functionally. If this is not possible, then the public agency needs to be cautious in interpreting the assessment results. The focus of this section of the law is to ensure that children are assessed fairly. The team must ensure that the evaluation is comprehensive and also sensitive to the child's impairment. For example, if a child has a reading disability and one of the assessments is of mathematics, then the assessment should be related to mathematics and not include the child's ability to read math story problems. Furthermore, these individual assessments must be conducted by trained individuals who are knowledgeable about the specific instruments and techniques used in this comprehensive evaluation. After the evaluation, if

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

a child is found eligible for services, then an IEP meeting is scheduled that includes the parents. The meeting is held and an IEP is written for the student that includes stated and measurable goals. At least once per year, the student's IEP is reviewed to examine progress toward those goals. As students age and mature, it is expected that there will be changes and a comprehensive reevaluation can occur after a year and should occur at least once every three years.

LO8 ***Family Educational Rights and Privacy Act of 1974.*** The next piece of legislation in this discussion also concerns education but includes both individuals with disabilities and those without. The Family Educational Rights and Privacy Act (FERPA) provides parents and "eligible students" (those older than 18) with certain rights regarding the inspection and

TABLE 5.2
Section 300.304 evaluation procedures from the Individuals with Disabilities Education Act of 2004

(a) Notice. The public agency must provide notice to the parents of a child with a disability, in accordance with Sec. 300.503, that describes any evaluation procedures the agency proposes to conduct.
(b) Conduct of evaluation. In conducting the evaluation, the public agency must—
  (1) Use a variety of assessment tools and strategies to gather relevant functional, developmental, and academic information about the child, including information provided by the parent, that may assist in determining—
    (i) Whether the child is a child with a disability under Sec. 300.8; and
    (ii) The content of the child's IEP, including information related to enabling the child to be involved in and progress in the general education curriculum (or for a preschool child, to participate in appropriate activities).
  (2) Not use any single measure or assessment as the sole criterion for determining whether a child is a child with a disability and for determining an appropriate educational program for the child; and
  (3) Use technically sound instruments that may assess the relative contribution of cognitive and behavioral factors, in addition to physical or developmental factors.
(c) Other evaluation procedures. Each public agency must ensure that—
  (1) Assessments and other evaluation materials used to assess a child under this part—
    (i) Are selected and administered so as not to be discriminatory on a racial or cultural basis;
    (ii) Are provided and administered in the child's native language or other mode of communication and in the form most likely to yield accurate information on what the child knows and can do academically, developmentally, and functionally, unless it is clearly not feasible to so provide or administer;
    (iii) Are used for the purposes for which the assessments or measures are valid and reliable;
    (iv) Are administered by trained and knowledgeable personnel; and
    (v) Are administered in accordance with any instructions provided by the producer of the assessments.
  (2) Assessments and other evaluation materials include those tailored to assess specific areas of educational need and not merely those that are designed to provide a single general intelligence quotient.
  (3) Assessments are selected and administered so as best to ensure that if an assessment is administered to a child with impaired sensory, manual, or speaking skills, the assessment results accurately reflect the child's aptitude or achievement level or whatever other factors the test purports to measure, rather than reflecting the child's impaired sensory, manual, or speaking skills (unless those skills are the factors that the test purports to measure).
  (4) The child is assessed in all areas related to the suspected disability, including, if appropriate, health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status, and motor abilities;
  (5) Assessments of children with disabilities who transfer from one public agency to another public agency in the same school year are coordinated with those children's prior and subsequent schools, as necessary and as expeditiously as possible, consistent with Sec. 300.301(d)(2) and (e), to ensure prompt completion of full evaluations.
  (6) In evaluating each child with a disability under Sec. 300.304 through 300.306, the evaluation is sufficiently comprehensive to identify all of the child's special education and related service needs, whether or not commonly linked to the disability category in which the child has been classified.
  (7) Assessment tools and strategies that provide relevant information that directly assists persons in determining the educational needs of the child are provided.

Authority: 20 U.S.C. 1414(b)(1)-(3), 1412(a)(6)(B)

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

dissemination of educational records. This law was designed to protect student privacy and applies to all school districts and schools that receive funds from the U.S. Department of Education. FERPA, which mainly concerns the access of educational records, provides parents with access to their children's educational record. Students over 18 also have access to their own records. The law further specifies that an educational record cannot be released, without parental permission, to anyone other than those who have a *legitimate* educational interest. The contents of the educational record will vary somewhat from school district to school district, but group achievement tests, student grades, and attendance records are routinely considered part of a student's educational record. Counseling records kept in a separate locked cabinet and accessible only to the counselor are not considered part of the educational record.

Another provision in FERPA concerns assessment in the schools. FERPA includes language stating that no student shall be required, without parental permission, to submit to psychological examination, testing, or treatment that may reveal information concerning mental and psychological problems potentially embarrassing to the student or the student's family. This issue was addressed in an amendment to the Goals 2000: Educate America Act of 1994. This amendment, which is titled the Protection of Pupil Rights Amendment and which is often called the Grassley Amendment, has language similar to FERPA, but it specifies written parental consent before minor students are required to participate in any survey, analysis, or evaluation that reveals any information concerning mental and psychological problems potentially embarrassing to the student or the student's family. Some educational administrators have interpreted this to include self-esteem measures or any exercise that involves self-analysis. Many school districts have adopted the policy that any type of assessment, except routine achievement tests, requires written parental consent before the assessment. Therefore, school counselors are encouraged to discuss with an administrator the school district's policy concerning any assessment they may want to initiate. This does not mean that counselors cannot use assessment tools, but in many instances, they are advised to secure written parental permission beforehand.

After some of the recent violent episodes at schools and universities, there has been debate about balancing students' privacy and school safety. In an emergency, FERPA does enable schools without consent to disclose education records that include personally identifiable information to protect the health or safety of students or other individuals. The release of information, however, is limited to the period of the emergency and it does not allow for unlimited release. Schools are increasingly using law enforcement units or security videos; these reports are not considered part of the student record and should be kept separately. Also, FERPA does not prohibit a school official from disclosing information about a student if the information is obtained through the school official's personal knowledge or observation, and not from the student's education records. For example, if a teacher overhears a student making threatening remarks to other students, this is not considered part of the student's educational record and the teacher may disclose what she overheard to appropriate authorities. If a student discloses to a counselor that he is going to harm himself or someone else, then the counselor is required to report this and students need to be aware of this exception in terms of confidentiality. If school personnel have questions about FERPA and releasing information, they can contact the U.S. Department of Education Family Policy Compliance Office.

LO9 ***Health Insurance Portability and Accountability Act of 1996.*** The Health Insurance Portability and Accountability Act (HIPAA) is applicable to counselors working in community agencies and requires that all clients be provided with notification describing how psychological and medical information may be used and disclosed as well as how the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

agencies can get access to that information. HIPAA requires the U.S. Department of Health and Human Services to establish national standards for electronic health care transactions and national identifiers for providers, health plans, and employers. These regulations apply to counselors who may be electronically transmitting mental health information. In addition, HIPAA concerns the security and privacy of health information that is sometimes related to psychological assessment information. HIPAA is quite complicated and beyond the scope of this book; furthermore, provisions of state law, which may be stricter about protecting a patient's health information, take precedence over related provisions of the HIPAA privacy rule. Although it is difficult to interpret the HIPAA regulations and identify which specific counseling agencies they apply to, many of the regulations are good practices. These practices include safeguarding clients' privacy and informing clients during the initial meeting of the agency's privacy practices. Clearly, if a counselor is involved with any third-party source (e.g., insurance company), then HIPAA regulations apply. According to Rada (2003), health care providers must supply a privacy notice to patients or clients, and health care providers must make a good-faith effort to obtain the patients' or clients' written acknowledgment of receipt of this notice. The notice should be in plain language and should outline the individual's privacy rights in a manner that clients will understand. The notice should describe the uses and disclosures of client information, include information on client rights, provide an overview of the counselor's responsibilities concerning privacy, indicate that clients will be informed if privacy policies change, and identify who clients can contact if they believe their privacy has been violated. Many counselors will specifically note exceptions to the authorized disclosures, such as child abuse, adult and domestic abuse, judicial or administrative proceedings, serious threats to health or safety, and worker's compensation situations. Pertinent to assessment results are the rights of clients, which include (1) the right to request restrictions on certain uses and disclosures, including a statement that the covered entity is not required to agree to a requested restriction; (2) the right to receive confidential communications of protected health information; (3) the right to inspect and copy protected health information; (4) the right to amend protected health information; and (5) the right to an accounting of disclosures of protected health information (Rada, 2003). Counselors are also responsible for developing, maintaining, and accounting disclosures of private client information that clients can access for a period of six years.

L10 Litigation

In the discussion of the second manner by which laws are established, litigation, the approach is somewhat different. Rather than discussing the information court case by court case, the focus is on areas within assessment where judicial decisions have been influential.

Test Bias and Placement. There have been some controversial decisions related to the use of intelligence tests for assigning African American children to for what was previously called classes for the mentally retarded (Fisher & Sorenson, 1996). The term *mentally retarded* is not used any more, but it was a common term when these court cases occurred. In the past in California, the classrooms for mentally retarded students had a substantial overrepresentation of African American students, and there were concerns that testing procedures were discriminatory. In the Federal District Court of San Francisco, Judge Peckham found in the case of *Larry P. v. Riles* (1979) that because intelligence tests discriminated against African American children, these tests could not be used to test for placement in educable mentally retarded classrooms in California. Less than a year later, Judge Grady in Illinois—who had a similar case, *Parents in Action on Special Education (PASE) v. Hannon* (1980)—ruled the opposite way, declaring that intelligence tests could be used when used in conjunction with other criteria. In both cases,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the rulings affected only those states. There was confusion in other states as well as about the legality of using intelligence tests with minority children. In a similar case in Georgia, *Georgia NAACP v. State of Georgia* (1985), the ruling was once again that intelligence tests did not discriminate. In 1986, Judge Peckham in California reissued his ban on intelligence tests. Therefore, in the mid-1980s, personnel in other states questioned the use of intelligence tests with minority students. However, in 1992, Judge Peckham lifted the ban on intelligence tests after African American parents sought to have their children assessed with intelligence tests to identify possible learning disabilities. Nevertheless, all of these cases resulted in the practice of using other criteria, along with intelligence tests, to determine whether a child has a mental disability.

Minimum Competency. During the 1980s, there was an upsurge of interest in competency testing or minimum competency testing. Considerable focus was on the lack of academic skills that some students had when graduating from high school. Fisher and Sorenson (1996) indicated that courts have consistently ruled that it is within the scope of the U.S. Constitution for states to require competency examinations for high school graduation. As with the example of intelligence tests, a judicial controversy related to discrimination also existed in the area of minimum competency examinations. Although other cases were filed in Florida related to minimum competency tests for high school graduation, the case of *Debra P. v. Turlington* (1981) was the most influential. A minimum competency examination was instituted in Florida in 1978, and a challenge was filed that year on behalf of all Florida seniors and on behalf of African American Florida seniors. A federal district court ruled that the exam violated all students' rights to procedural due process and violated African American students' right to equal protection. The court postponed the use of the examination until the 1982–1983 school year. The reason for the postponement was that Florida schools were segregated until 1971, and the court ruled that the African American students had not received an equal education. Starting in the 1982–1983 school year, all seniors would have received all of their education in desegregated schools and, thus, had equal educational opportunities. This case is also important because it raised the question of instructional validity. Instructional validity concerns the evidence that students were taught the material covered on the test. The court ruled that if the test covers material not taught to the students, then the test is unfair and violates the equal protection and due process clauses of the U.S. Constitution. This ruling set a precedent concerning the need for a relationship between the curriculum and the minimum competency examination in each state.

In a related case, *GI Forum v. Texas Education Agency* (1997), the plaintiffs argued that the Texas graduation test unfairly discriminates against minority students because the percentage of Hispanic and African American students passing this test was lower than white students. In particular, they asserted that tests administered to nonnative English speakers should be designed to minimize threats to reliability and validity related to English language proficiency. In reviewing the evidence from multiple sources including the *Debra P.* case, the judge in *GI Forum v. Texas Education Agency* found that the graduation test was reliable and did assess the state-mandated curriculum. The judge in this case also considered court cases related to employment testing. For employment testing, a guideline that has been supported in the court is that adverse impact occurs when the passing rate for the minority groups is less than 80% of the passing rate for the majority group. In this case, there were a few years in which the percentage of Hispanic and African Americans was less than 80% of whites, but the court ruled that cumulative records can be used in determining if a graduation test is having an adverse impact. This case is noteworthy because it supported the 80% rule in educational testing. For more information on this case, interested readers are directed to Phillips (2000).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Summary

This chapter discussed ethical and legal issues in assessment. Ethics concerns the principles that guide professionals in their performance of duties. Laws are legal rules enforced by governmental entities. A number of resources, including the American Counseling Association's (2014) *Code of Ethics* and the *Standards for Educational and Psychological Testing*, are available to assist counselors when faced with ethical dilemmas regarding assessment. In terms of ethical responsibility, a clinician bears the ultimate responsibility for the appropriate use of any instrument. Although test publishers may attempt to restrict the sale of instruments to untrained individuals, ethical standards clearly indicate that a practitioner must practice within his or her level of competency. Because instruments vary in the amount of training and experience that is necessary in order to be competent, counselors need to ensure that they have been adequately trained before giving an instrument to a client.

Laws related to assessment have emerged from both legislation and litigation. Concerning legislation, the Civil Rights Act of 1991, the American with Disabilities Act Amendment Acts, and the Individuals with Disabilities Education Act have all influenced the practice of assessment. The crux of all of these laws is the use of sound and appropriate instruments with individuals in a manner that is fair and equitable. The Family Educational Rights and Privacy Act governs who has access to educational records and assessments in the schools. The Health Insurance Portability and Accountability Act (HIPAA) also governs the privacy of health information, which could include psychological assessments in some counseling settings. Litigation has primarily involved test bias and minimum competency examinations. It is important for counselors to stay informed about pending legislation and litigation that have the potential to affect assessment in counseling. Because social factors influence litigation and legislation, there will be changes in the laws governing appraisal as our society changes in the coming years.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 6

Issues Related to Assessment with Diverse Populations

Learning Objectives

- LO1 Describe accessibility and universal design
- LO2 Distinguish among adaptation, accommodation, and modification
- LO3 Understand how to assess an instrument for content bias
- LO4 Comprehend how to investigate the internal structure of an instrument
- LO5 Describe how to evaluate an instrument's relationship with the criterion
- LO6 Explain the major points in the *Standards for Multicultural Assessment*
- LO7 Discuss differences in performance based on race and/or ethnicity
- LO8 Examine the issues related to assessing nonnative English speakers
- LO9 Comprehend the issues related to the assessment of individuals with disabilities

You are working as an elementary school counselor. A kindergarten teacher has requested your advice about a student named Maria, who immigrated with her parents to the United States from Costa Rica within the past month. Maria does not speak any English. In this small school district, the kindergarten teachers do an informal kindergarten readiness assessment, but Maria's teacher only feels comfortable in conducting these assessments in English. She did not assess Maria, but she is concerned about whether Maria is ready for kindergarten because Maria's birthday is only two weeks before the cutoff date for matriculation. Furthermore, Maria whimpers through most of the school day and does not interact with any of the children on the playground. The teacher wants your advice on how to assess Maria's level of academic and emotional development.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In assessment, there is the assumption that everyone has an equal opportunity to perform. The assumption may not always be correct, for there may be a variety of reasons that an assessment may be fair to some individuals and unfair to others. Individuals' backgrounds may vary, and they may have different experiences and knowledge to draw from while being assessed. For example, let's say that only people who could sing the Indiana University fight song (i.e., *Indiana, Our Indiana*) could be hired as counselors in a counseling agency. Many counselors would be able to learn this song but were probably not exposed to it during their counseling classes. When it comes to assessment, questions related to equal opportunity to learn or to be exposed to material can often be an issue. In the case of the Indiana University fight song being used as an employment test, it might be slightly unfair because those who attended Indiana University may have an advantage. In many testing situations, the influences of background and culture are not as blatant as in the Indiana University example. Culture has a significant influence on all aspects of the individual, and problems can occur with instruments that do not consider cultural differences. In pondering the ludicrous example of the fight song, we might also question its fairness for individuals who, for some reason, are physically unable to sing. Once again, although this example is somewhat asinine, it is raised in the hopes that readers of this book will begin to reflect on the importance of considering background, opportunities, barriers, and physical and emotional disabilities in the area of assessment.

Knowledge of using appropriate assessment techniques with diverse populations is crucial. One of the reasons for updating the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) was to focus more on test fairness, which is infused throughout the *Standards*, and an entire chapter is devoted to the issues of fairness in testing. Furthermore, many counseling-related ethical standards (e.g., those of the American Counseling Association, the American Psychological Association, and the American Association of Marriage and Family Therapists) contain specific information about assessment with special populations. According to the *Standards*, fairness is interpreted as responsiveness to individual characteristics and testing contexts so that instrument scores will result in valid interpretations. Therefore, fairness is fundamentally an issue of validity and it requires attention throughout all stages of instrument development and use.

In addressing issues of fairness in assessment, one needs to return to the topic of construct-irrelevant variance, which was discussed in Chapter 4, and its focus on validation evidence. A counselor wants an instrument that measures the same construct fairly for all individuals and one that does not measure construct-irrelevant attributes for certain individuals. According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), "the central idea of fairness in testing is to identify and remove construct irrelevant barriers to maximal performance for any examinee" (p. 63). An example of construct irrelevance is an assessment that is attempting to measure the construct of mathematical knowledge; however, many of the items are story problems so that individuals' reading abilities are irrelevant to the measurement of mathematical knowledge.

A very helpful resource in the area of assessment with diverse population is the Association for Assessment in Counseling and Education's (AACE) *Standards for Multicultural Assessment* (AACE, 2012), which is included in Appendix C. The AACE is now known as the Association for Assessment and Research in Counseling (AARC). The *Standards for Multicultural Assessment* has been revised four times, and this recent revision addresses more specifically the role of social advocacy in assessment. They define multicultural populations broadly to include individuals who differ by race, ethnicity, culture, language, age, gender, sexual orientation, religion, and ability. This is consistent

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

with my view of multiculturalism, and readers are encouraged to include this broad definition while reading this chapter.

LO1 Before talking about the traditional ways of evaluating instruments for biases, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) introduces two relatively new concepts in considering instruments for diverse populations and fairness: **accessibility** and **universal design**. The first concept, accessibility, concerns the notion that all examinees should have an unobstructed opportunity to demonstrate their capabilities on the construct(s) being assessed. A good example of the concept of accessibility is an individual with limited English proficiency who may not be adequately diagnosed if the structured interview is conducted in English. Another example of accessibility is an older client with a hearing impairment who may not adequately hear the questions in this same structured clinical interview. As these examples demonstrate, access to the construct being measured can be obstructed and this influences that validity of the interpretation.

The second relatively new concept in this area is that of universal design. Universal design is an approach to instrument design with the goal of maximizing accessibility for all intended examinees (AERA, APA, & NCME, 2014). With universal design, the intent is to develop and design instruments from the beginning that allows for the participation of the widest range of individuals rather than retrofitting existing instruments through accommodation or a variety of alternate assessments (e.g., instruments in different languages). With universal design, the developers begin the process of developing an instrument with the idea of maximizing fairness. The first step is to define the construct to be measured precisely in order to minimize the influences of construct-irrelevant factors. The next step is to write accessible, nonbiased items. In this case, the font size may be increased so it is easier for those with visual impairments to read or the timing of items might be relaxed when it is not relevant to the construct being measured. Also, item formatting is adopted that is familiar to the largest number of examinees with an attempt to minimize formatting that is unfamiliar to some because of their cultural background. The instructions and procedures for the instrument are developed to be clear, concise, and intuitive. For more information on universal design, please see Lovett and Lewandowski (2015).

LO2 Although most instrument developers want to design assessment instruments that are fair to all, that is not an easy enterprise and sometimes it is not possible. This leads to issues of adaptation, accommodation, and modification. **Adaptation** is changes to the original instrument in terms of design or administration to increase the accessibility to individuals (e.g., visually impaired, limited English proficiency). With adaptation the instrument may be converted to a Braille format or translated into a second language. Although sometimes the terms *adaptation*, *accommodation*, and *modification* will be used interchangeably, **accommodation** typically represents a change in administration where the comparability of the scores is retained (AERA, APA, & NCME, 2014). An example of an accommodation is allowing more time to take a test for a person who has been diagnosed with attention-deficit/hyperactivity disorder (ADHD). With **modification**, the changes being made influence the construct being measured and, therefore, the scores do not retain the same meaning as the original instrument.

Culture and Assessment

Individuals from different cultures vary in many aspects, including how they view the world. According to Dana (2005), *worldview* involves group identity, individual identity, beliefs, values, and language. Worldview influences all perceptions and, thus, influences not only what is learned but also how people learn. Therefore, clients' culture and worldview can affect their performance on traditional achievement, aptitude, and ability tests. Assessment in the affective domain may also be influenced because clients' worldviews influence responses to questions, nonverbal behaviors, and communication styles. To work effectively with diverse clients, counselors need to understand differences in worldview.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Counselors should also consider their own worldview, which is sometimes conceptualized from the etic versus emic perspective. An etic view is a *universal* perspective where an assessment is believed to measure a universal trait or construct. With an emic perspective, culture is always considered in assessment. As an example, with diagnosis from an etic perspective, the individual would believe that the major features of the disorder will manifest itself regardless of culture. From an emic perspective, however, the clinician would understand the relation of culture in the manifestation of a possible disorder (Suzuki, Alexander, Lin, & Duffy, 2006).

Culture is defined as the belief systems and value orientation, including customs, norms, practices, and social institutions (Fiske, Kitayama, Markus, & Nisbett, 1998). A postmodern perspective would suggest that language has a pervasive influence affecting perceptions of reality. The influence of language needs to be considered in the assessment techniques most commonly used by counselors, which are primarily verbal techniques such as questioning, clarifying, and reflecting. For example, cultures differ in terms of the degree to which self-disclosure, particularly to a nonfamily member, is encouraged. An insensitive counselor may perceive the client's reticence to open up as resistance rather than seeing that it is due to cultural differences. This perception could encourage misunderstandings and problems in the client-counselor relationship. Because counseling is predominantly a verbal process, the subtle influences of culture on language need to be considered by a multiculturally competent practitioner.

Types of Instrument Bias

As stated earlier, in assessment it is assumed that all individuals are given an equal opportunity to demonstrate their capabilities or level of functioning during the assessment process. However, even an instrument designed to be culturally sensitive and fair may be unfair, depending on the circumstances under which it is used or the groups of individuals with whom it is used. The term **bias testing** refers to the degree that construct-irrelevant factors systematically affect a group's performance. In other words, factors not related to the instrument's purpose either positively or negatively influence a group's performance. Given the potential consequences of using a biased instrument, it is important for counselors to know how to evaluate instruments for the various types of possible bias.

LO3 ***Content Bias.*** Padilla and Borsato (2008) contended that appropriate multicultural assessment starts at the very beginning, when an instrument is just being initiated. Instruments may be biased: the content may be more familiar or appropriate for one group as compared with another group. As an example, let's imagine that an intelligence test has a question regarding what the state bird of Wyoming is. Many individuals would have difficulty remembering this piece of trivia; however, if you had attended school in Wyoming, you could probably answer that the Wyoming state bird is the meadowlark. Thus, individuals from Wyoming would have an advantage on this intelligence test over individuals from other states. The biased content of an instrument may be subtler than the Wyoming state bird question. For example, some groups could perceive certain items on an instrument as irrelevant or not engaging. Children living below the poverty level may see little relevance in a question concerning how long it will take an airplane to fly from Point A to Point B when these children cannot imagine ever having the financial resources to fly on an airplane. Children may be less likely to engage in questions when the names and circumstances on the test are always associated with the majority culture. Bracken and Barona (1991) found evidence to suggest that children from different cultural backgrounds interpret test items differently. In considering culture in assessment, Padilla and Borsato (2008) maintained that issues of English-language proficiency, age of immigration, length of time in the United States, ethnic identity, and extent of acculturation are often not considered in the test development phase.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In evaluating an instrument for content bias, the practitioner needs to examine the procedures used in developing the instrument. Haphazard selection of instrument content may result in an instrument with content bias. Specific attention needs to be paid to multicultural issues during the initial stages of instrument development. In the manual, instrument developers should document the steps they took to guard against content bias. Care in writing items that appear to be nonbiased may not be sufficient. A common procedure is to have a panel made up of diverse individuals review the instrument. Nevertheless, content bias can be difficult to uncover because subtle biases may not always appear by simply examining the instrument.

In addition, statistical methods are often employed to supplement review of the content. One statistical method for investigating item bias that has received increasing attention is **differential item functioning (DIF)**, which was briefly mentioned in Chapter 4. DIF is a method for analyzing the relative difficulty of individual items for different racial or ethnic groups. Although there are different methods for performing DIF, all of them strive to examine differences in performance among individuals who are equal in ability but from different groups. If an item is found to have unequal probabilities for correct responses from certain groups with equal ability, then the item may be biased. Sometimes, however, statistical artifacts can result in group differences rather than the item being biased. In addition, more sophisticated models of DIF involve matching examinees across groups to control for group differences (Sireci & Geisinger, 1998). Although DIF techniques have been very useful in identifying and eliminating individual items on specific instruments, they have not identified common themes or characteristics of items that are prone to bias. Advances are being made in this area, and we hope that research will soon shed light on methods for developing unbiased items.

LO4 ***Internal Structure.*** Another method for detecting possible instrument bias is to examine the instrument's internal structure. For instance, an instrument's reliability coefficients for one ethnic group could be very high, whereas for another ethnic group, they could be significantly lower. Thus, an instrument can be a reliable measure for one client but not for another. Well-developed instruments will investigate this problem and, in the manual, present reliability information for different groups. Typically, researchers investigate differences between males and females and between different ethnic groups. The investigation of differences in reliability among different ethnic groups depends on the size and representation of the norming group. Hence, if there were problems with the makeup of the norming group in terms of minority participants, then it would be difficult to analyze differences in reliability.

Examination of internal structure as related to instrument bias can also involve the instrument's factor structure. As discussed in this book, many instruments and tests used in counseling have subscales. It is important for researchers to determine if the dimensions or factors underlying the instrument vary for different groups. For example, in personality instruments, the factor structure for an ethnic group may not match the subscales of the instrument; thus, it would not be appropriate to interpret those subscales in the typical manner. Practitioners need to examine the research related to the factor structure of an instrument with subscales to determine if it is appropriate to interpret the scores of the subscales with clients from diverse backgrounds.

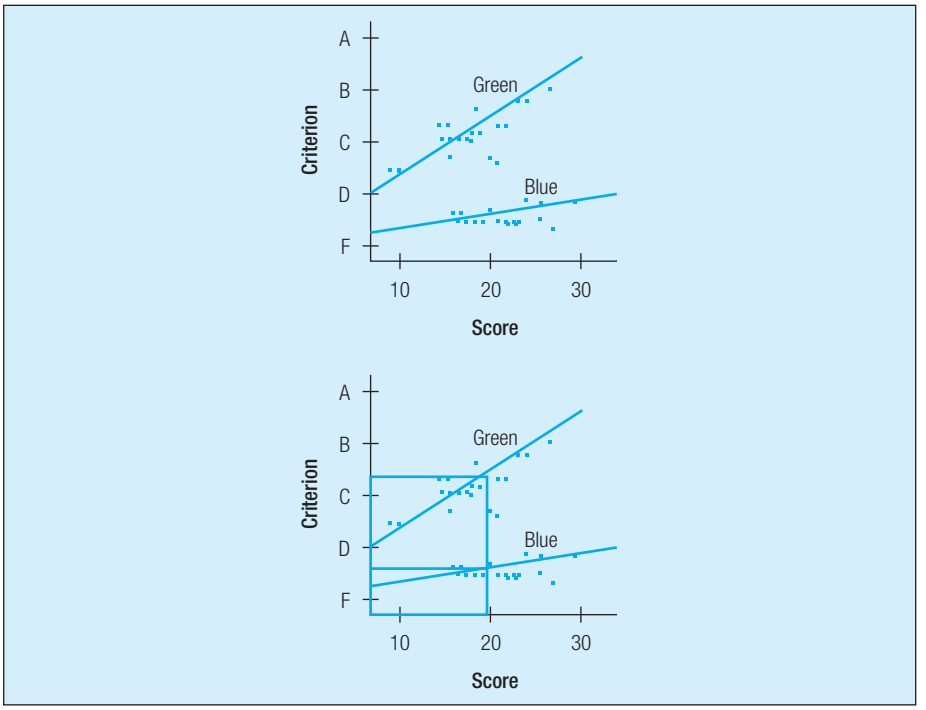
LO5 ***Instrument and Criterion Relationships.*** The analyses of possible bias should also include studies related to the relationship between the instrument and the criterion for different groups. For example, an instrument may be highly related to job performance for one ethnic group but may have a low relationship for another ethnic group. It is important to examine the differences among the validity coefficients, not simply whether the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

validity coefficients are statistically significant for one group and not significant for another group. Statistical significance is influenced by sample size; typically, the sample size is large for whites or European Americans and small for African Americans. With correlation, the smaller the sample, the larger the correlation coefficient needs to be in order to be significant. Therefore, the exact same validity coefficients for these two groups could be significant for European Americans and not significant for African Americans. Most of the research related to ethnic differences in validity coefficients has compared African Americans and European Americans. Very few studies, however, have examined the differences in validity coefficients for Hispanics and other ethnic groups.

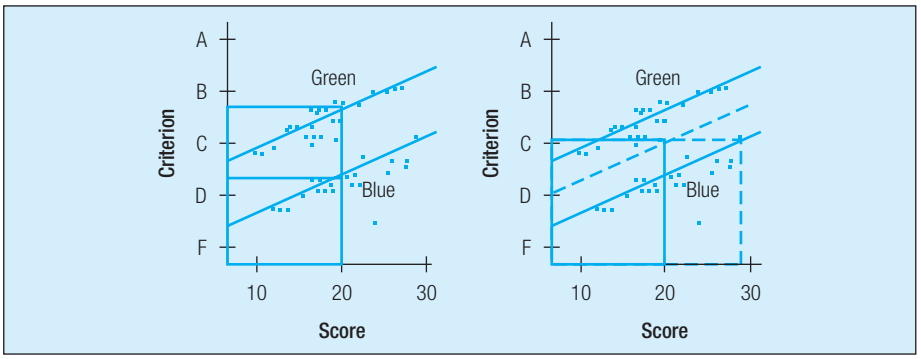
Significant differences in validity coefficients can have a notable effect if the instrument is used for selection or predictive purposes. In fact, the *Standards for Educational and Psychological Testing* (2014) clearly state that "correlation coefficients provide inadequate evidence for or against a differential prediction hypothesis if groups are found to have unequal means and variance on the test and the criterion" (p.52). The statistical technique of regression is often used for selection, and when there are different validity coefficients, the instrument can vary in how well it predicts to the criterion for the different groups. If an instrument yields validity coefficients that are significantly different for two or more groups, then this difference is described as **slope bias**. Figure 6.1 reflects slope bias because, as you can see, the same score on the instrument would predict *different* performance on the criterion depending on whether a client is in the green group or the blue group. In this example, a score of 20 by a green individual would predict a grade of a little over a C, whereas for a blue individual a score of 20 would predict a grade between D and F. In cases of slope bias, some experts in the field of assessment have suggested using separate regression equations for the different groups. As discussed in Chapter 5, however, there are certain situations in which legislation has prohibited the use of separate regression equations for different ethnic groups.

FIGURE 6.1
Slope bias



Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 6.2
Intercept bias



Even instruments with the same validity coefficient can predict different criterion scores for members of different groups with the *same score*. If you examine Figure 6.2, the regression lines for the green and the blue groups are parallel, which indicates that the validity coefficients are the same. The lines, however, differ in where they begin on the y-axis; hence, the **intercepts** are different. The first graph in Figure 6.2 indicates that a score of 20 on the instrument predicts a grade slightly above D for an individual from the blue group and that same score predicts a grade of almost a B for someone from the green group.

Thus, using a cutoff score of 20 provides an advantage to the blue group. As the second graph in Figure 6.2 reflects, in using the intercept for the blue group an individual from this group would need a score of almost 30 to be predicted to receive a grade of C. Numerous studies have found that there are no intercept differences among ethnic groups on many of the instruments in counseling (Anastasi & Urbina, 1997). Some researchers have found that when there are differences among majority and minority groups, majority group members tend to have higher intercepts than do minority members. Majority groups have also been found to have higher intercepts for some instruments that predict college grades (Duran, 1983) and job performance (Hunter, Schmidt, & Hunter, 1979). Thus, in Figure 6.2, the majority group members would be the green group and the minority members would be the blue group. Therefore, these instruments would overpredict, rather than underpredict, minority individual performance. Linn (1994), however, argued that evaluating instruments based solely on regression models is inadequate and that several artifacts of measurement affect this overprediction for certain ethnic groups.

This section of the chapter addressed methods for examining whether an instrument is biased. It is possible for an assessment instrument to have culturally sensitive content, almost identical validity coefficients for ethnic groups, and the same intercepts yet produce very different scores for individuals from different ethnic groups. In the discussion of ability, achievement, and aptitude testing, minority individuals tend to have average scores lower than those of majority individuals on many of the instruments. Even though some of these instruments may not be biased, the result of using these instruments has a disparate or negative impact on some minority individuals.

L06 General Recommendations for Practice

Counselors can expect to be involved in either formal or informal assessment of clients from cultures other than their own. Padilla and Borsato (2008) suggested that practitioners may not be aware of how their linguistic, social class, and cultural experiences influence their construction of knowledge, selection of particular tests, and interpretation derived from assessment procedures. Thus, counselors need to find methods for effective cross-cultural assessment. As indicated earlier, one excellent resource for counselors is the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Standards for Multicultural Assessment (AACE, 2012). The following discussion centers on each of the standards from the *Standards for Multicultural Assessment*.

Culturally competent professional counselors recognize the importance of social justice advocacy; they integrate understanding of age, gender, ability, race, ethnic group, national origin, religion, sexual orientation, linguistic background, and other personal characteristics in order to provide appropriate assessment and diagnostic techniques.

As mentioned earlier, the most recent version of the *Standards for Multicultural Assessment* focuses on social justice advocacy. In this first standard, AACE (2012) emphasizes that counselors recognize in themselves and others subtle biases and further explore how these biases influence the assessment process for marginalized populations. These standards further encourage counselors to learn by immersion about marginalized population in order to understand clients' worldviews and understand the impact of this worldview on the assessment process. Counselors are encouraged to create awareness about assessment with diverse clients and support the use of assessments that are psychometrically appropriate for clients from diverse groups. Furthermore, counselors are urged to provide culturally competent and effective practices in all areas of counseling and assessment and to work collaboratively with community leaders to address the needs of diverse clients. More specifically, counselors are to address systemic barriers and consider how these barriers might impact the use and interpretation of assessment results. Counselors are also encouraged to be knowledgeable about potential biases in assessment instruments and comply with ethical guidelines when assessing marginalized groups. Moreover, counselors are responsible for the appropriate use of assessment instruments whether they score and interpret such assessment or use technology or other services. Finally, counselors must take reasonable steps to ensure the proper use of psychological and assessment techniques by person under their supervision.

Culturally competent professional counselors select assessments and diagnostic techniques that are appropriate and effective for diverse client populations.

When developing a new instrument, counselors should ensure the technical quality of the instrument, and understand and follow standard development procedures that utilize accepted research methods (AACE, 2012). As is emphasized in this book, counselors should also be knowledgeable of the technical aspects of assessment, which includes reliability, validity, measurement error, and appropriate norms. Before selecting an instrument, counselors are also encouraged to evaluate the directions and a sample of test questions or practice tests, answer sheets, manuals, and score reports. Counselors must be able to understand how to review assessment information regarding diverse subgroups, which includes using appropriate norming information to ensure adequate subgroup analyses. For clients with disabilities who need special accommodation, counselor should understand how to select and utilize appropriate modified forms of the instrument. In order to facilitate client empowerment and self-advocacy, counselor should select assessments that assist in identifying client's needs, strengths, and resources. Counselors only use the instruments in which they are competent to use and ensure that persons under their supervision are also aware of these multicultural standards. Finally, counselors recognize the impact of cultural identity on instrument administration and interpretation, and consider assessment results with other relevant factors.

Culturally competent professional counselors recognize challenges inherent in assessment of persons and seek to provide administration and scoring of assessment to clients respecting age, gender, ability, race, ethnic group, national origin, religion, sexual orientation, linguistic background, and other personal characteristics.

The first standard in this section addresses the topic of utilizing a competent and bilingual translator, and how even with a competent bilingual translator, counselors must

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

obtain, understand, and follow all established protocol for administering standardized test instruments to clients with diverse linguistic backgrounds. Therefore, a counselor cannot simply rely on a translator to administer an assessment, but the counselor must ensure that he or she is working with a competent translator and follow the assessment protocol during the administration of the standardized assessment. Furthermore, the counselor should be familiar with the formatting of the test question, procedures for answering test questions, and needed equipment or additional materials that are necessary in the administration of the instrument. Counselors also need to understand the importance of test security, copyright violations, and the ramifications of cheating. They further keep scores secure and maintain the confidentiality of scoring results. Also of importance for counselors is knowledge of the directions, procedures, and materials for scoring assessments and/or directions for monitoring the administration of the assessment. If an error occurs in the administration, then in many circumstances counselors must promptly report this error. Counselors also only administer assessments within the context of a defined counseling relationship and only use assessments for their recommended purposes. Furthermore, counselors should obtain informed consent from clients regarding the nature and purpose of assessment and ensure that the client has the opportunity to become familiar with the formatting of the questions (AACE, 2012).

Culturally competent professional counselors acknowledge the importance of social justice advocacy in interpretation and communication of assessment results with diverse populations.

In both scoring and interpretation, the counselor must understand the potential for bias and be knowledgeable of the norming group, technical information, and benefits and limitation of the assessment (AACE, 2012). Counselors are further encouraged to provide a holistic appraisal of clients by creating a strong foundational evaluation that respects the influences of culture, background, and individual characteristics. This understanding of the influences of culture, background, and individual characteristics are also used when designing and implementing interventions to achieve effective counseling outcomes. Also for culturally diverse clients, counselors recognize the effects of stigma, oppression, and discrimination on the possible interpretation of assessment results and how those results may be used. Counselors also recognize and collaborate with others to eliminate biases, prejudices, and discrimination in the context of assessment, interpreting results, and providing interventions. Moreover, counselors explain to the client or the client's legally authorized representative the nature, purpose of the assessment, and specific use of results in a manner that is understandable and developmentally appropriate. The counselor includes in this explanation the impact of culture on the assessment results. When relevant, the counselor also considers other client factors, such as culture, disability status, systemic oppression, or internalized oppression, before making recommendations. When releasing assessment data where the client is identified, the counselor only does so with permission from the client or his or her legal representative. Furthermore, the counselor only releases assessment information to other professionals who are recognized as qualified to interpret the assessment results.

Culturally competent professional counselors seek training and supervised experience to ensure they provide appropriate assessment and diagnostic techniques for diverse client populations.

In regard to seeking training and supervision, counselors seek guidance about the level of expertise needed to use an instrument and understand the characteristics of an instrument before using it. Furthermore, counselors understand individual and group approaches to multicultural assessment. Counselors also receive training related to how assessment data can be used to make systemic changes that benefit disadvantaged individuals and vulnerable groups. In order to develop assessment practices that empower clients and

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

inform the general public about culturally appropriate assessment, counselors collaborate with allies and policy makers. Counselors also engage in continuous education to increase their knowledge of assessment and enhance job performance. Finally counselors need to have training and expertise in traditional testing instruments and assessment including technical aspects of instruments as well as cultural limitations of instruments.

LO7 Differences in Performance and Race/Ethnicity

Today, counselors need to be prepared to work with clients from diverse racial and ethnic backgrounds. According to the U.S. Census for 2010, 72.4% of the population is white alone; 12.8% black or African American alone; 0.9% American Indian or Native Alaskan alone; 4.8% Asian alone; and 16.3% Hispanic or Latino. Interestingly, more than 50% of the growth in the total population of the United States between 2000 and 2010 was a result of the growing Hispanic population (U.S. Census Bureau, 2011). Projections indicate that the United States will become increasingly diverse; hence, counselors need to consider diversity issues within assessment. Before beginning a discussion of racial differences in assessment, I need to acknowledge some of the thorny issues about discussing race. According to Helms, Jernigan, and Mascher (2005), racial groups or categories are not psychological constructs because they do not automatically result in explicit behaviors, traits, biological factors, or environmental conditions. As an example, someone who is depressed acts depressed, but someone who is African American does not necessarily act African American. Helms et al. contended that racial categories are sociopolitical constructions that society uses to aggregate people on the basis of ostensible biological characteristics. There is substantial research that African Americans score almost a standard deviation below European Americans in intelligence or cognitive ability testing (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). The average scores for Hispanics or Latinos on intelligence tests tend to fall between the average scores for African Americans and European Americans (Suzuki, Short, Pieterse, & Kugler, 2001). A pertinent issue concerns what conclusions can be drawn from these findings regarding racial group performance on assessments. As indicated in Chapter 9, cognitive ability or intelligence assessment is not tapping some biological construct. Furthermore, as Helms et al. asserted, in counseling and psychological assessment, ascribed racial categories are typically associated with group-level socialization experiences that may or may not apply to the individual. For example, not all African American students attend inner-city schools with low achievement scores. Helms et al. (2005) contended that in counseling we should stop using race to categorize assessment results and use measures of constructs such as racial identity that have been found to explain significant differences in assessment of cognitive abilities and other areas of psychological assessment. Some measures of racial identity are sensitive to issues of oppression and privilege and can provide a better understanding of why individuals may perform differently on various assessments.

Although there are problems associated with using race to categorize people, counselors cannot ignore that currently in our society there are racial differences in average performance on many psychological assessments. The impact of these differences is particularly noteworthy when we examine how assessments of cognitive ability, knowledge, and skills are currently being used. For example, there is substantial research that has found the overrepresentation of minority students in special education (National Research Council, 2002; Zhang, Katsiyannis, Ju, & Roberts, 2014). Moreover, Skiba, Poloni-Staudinger, Gallini, Simmons, and Feggins-Azziz (2006) found that African American students diagnosed with emotional disturbances, mild mental retardation, moderate mental retardation, learning disabilities, and speech and language difficulties were more likely to be placed in more restrictive educational environments than other students with comparable issues.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

There are various opinions as to the causes for racial differences in test performance. The term *test* is used in this context because most of the debate has concerned differences in ability, achievement, and aptitude testing. It should be noted, however, that occasionally there are ethnic differences on some personality instruments. Some have argued that these differences are primarily innate and that little can be done to remedy this situation (Jensen, 1972, 1980). Others have argued that environmental factors are the major reason for the discrepant scores among racial groups and that unequal access to educational and stimulating experiences influences test performance. Those arguing for the latter perspective contend that tests highlight the unfairness of life, not the unfairness of tests.

For a time, some professionals in assessment focused on developing *culture-free* assessments. In general, experts in the field of assessment have concluded that, as of yet, there is *no* culture-free assessment device. Even with nonverbal tests, such as putting puzzles together or deciphering matrices, there are cultural components that influence performance. Because it has been impossible to develop an instrument devoid of cultural influences, attempts have been made to develop *culture-fair* instruments. Within the field of assessment, some individuals contend that culturally fair instruments already exist. Others, however, suggest that an instrument has yet to be developed that is culturally fair to all individuals. Stobart (2005) cogently argued that fairness in assessment cannot be separated from fairness in access opportunities and curriculum. He asserted that large-scale achievement tests highlight the sociocultural inequities in the U.S. educational system and that testing should be used to examine the opportunities and curriculum of those students whose performance is weak. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) contended that opportunity to learn is a fairness issue and in high-stakes testing, there must be evidence that students have been provided instruction and curriculum related to the constructs being assessed.

Helms (2004) defined *fairness* as the removal from test scores systematic variance that is related to the test takers' psychological characteristics that were influenced by socialization practices or environmental conditions "irrelevant to measurement of the construct of interest to the test user or assessor" (p. 87). Hence, Helms (2006) contended that test fairness or unfairness is often not due to the structure of the test but to the interaction among the examinees' internalized racial or cultural experiences, their environmental socialization, and the testing process. Consistent with how the profession now views reliability and validity, Helms argued that test fairness is not something that an instrument has or does not have from sample to sample. Her view of fairness in assessment is significant in that the counselor needs to consider the client's psychological characteristics and internalized racial or cultural experiences in interpreting test results. Consequently, individuals may have some similar experiences in life but how those experiences are internalized may be very different. Therefore, in testing, there may be construct-irrelevant variance for some individuals but not for others.

In discussing racial test score disparity, Helms (2002) presented the results from a study that examined her 1990 immersion-emersion racial identity scheme (e.g., idealization of one's blackness and black culture). Helms found that the difference between those not having any black culture and those having an average amount of black culture was approximately 129 points on the SAT-Verbal. Interestingly, this difference of approximately 1 standard deviation is very similar to the difference typically found between white and minority student scores on the SAT. In a tongue-and-cheek manner, Helms suggested that these results mean that the remedy for improving black test scores is either to deliberately destroy positive racial identity development of black students or to socialize white students to idealize black identity, thus eliminating the automatic increase of 1 standard deviation of white cultural advantage per subtest.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Helms's (2002, 2006) writing reinforces the importance of not considering instruments as being either valid or invalid because an instrument may be valid with one group but may have a very different meaning for individuals from another culture. In a discussion of multicultural assessment, it is also valuable to note the significance of socioeconomic status and the importance of considering socioeconomic background. Socioeconomic status is particularly critical because some researchers have found that it has more influence on test performance than either ethnicity or instrument bias (Groth-Marnat, 2009a; Liu & Watt, 2013). Researchers have found that differences on intelligence tests (Loehlin, 1989) and on the MMPI (Timbrook & Graham, 1994) dissipate when African Americans and European Americans of equal socioeconomic status are compared.

An important concept to consider in interpreting assessment results with racially and ethnically diverse individual is **acculturation**. Acculturation is generally viewed as "a dynamic process of change and adaption that individuals undergo as a result of contact with members of different cultures. This change is influenced by the environment the individual resides in, as well as qualities innate to that individual. These changes influence attitudes, beliefs, values, affect, and behaviors of those undergoing the acculturation process" (Rivera, 2008, p. 76). At first, acculturation was considered a one-dimensional concept, which occurs on a continuum ending with assimilation into mainstream culture. More recently, scholars acknowledge that it is a two-dimensional process that involves both the culture of origin and the host culture. Within this two-dimensional model, the individual is in the process of retaining parts of the culture of origin while simultaneously adapting to the new or host culture. Rivera (2008) contended that the complexities of measuring and considering acculturation can be daunting. Although there are numerous measures of acculturation, many clinicians use informal methods for assessing it. In coalescing models of qualitative assessments of acculturation, Ponterotto, Gretchen, and Chauhan (2001) suggested assessing clients along four dimensions: (1) client's worldview and perception of the problem, (2) client's family background, (3) cultural explanations of the individual's presenting illness, and (4) cultural elements of the counselor-client relationship.

LO8 Linguistic Background and Appraisal

Issues related to cross-cultural assessment are already complex, but the issues increase in complexity when language and nonnative English speakers are considered in the assessment process. In 2011, 22% of school-aged children spoke a language at home other than English (Childstats.gov, 2013). Language skills almost always have some degree of influence on a person's performance on an assessment device. For example, reading or listening skills are typically needed to hear or read test instructions. Even with arithmetic or mathematical assessment, most tests require an examinee to read either the instructions or some items. There are also differences in terms of language proficiency, and these differences may introduce construct-irrelevant factors into the assessment process. For some examinees, their results on a test may be more a reflection of their proficiency with English than of their ability in mathematics. These linguistic factors become even more crucial when important decisions are made based on assessment results. Therefore, counselors need to be aware of concerns related to assessing those with limited proficiency in the dominant language. It should be noted that many of these issues pertain not only to nonnative speakers but also to those with unique linguistic backgrounds, such as clients who are deaf or those who have other disabilities that affect the development of language.

Counselors are responsible for ensuring that all examinees are given the same opportunities to demonstrate their level of competence on the attributes being assessed. In our culture, when English is a client's second language, the counselor needs to consider numerous factors to ensure fair assessment. One of the factors to consider is the purpose of the testing

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

or assessment. If the purpose is to evaluate English proficiency, then there may be little need for adaptation of the testing situation. For example, if successful job performance in a position requires the ability to communicate in English (e.g., a technical writer), then it may be legitimate to use a test that assesses the ability to write in English. There are, however, a number of occupations for which the ability to communicate in English is not required. It is very rare that a clear distinction can be made concerning the influences of language and how to assess clients appropriately. Take, for example, the identification of a learning disability with a child whose dominant language is Spanish. It is sometimes difficult to decipher when the assessment results are related to language issues and when there is evidence of a learning disability. Assessment of individuals in a situation where there is a question of language proficiency requires special attention to administration, scoring, and interpretation.

In education, there are some legal issues related to assessment of English-language learners. According to the U.S. Supreme Court decision in *Lau v. Nichols*, providing students with the same resources (e.g., textbooks), teachers, and curriculum does not ensure equal educational opportunity for English-language learners. School districts are required to identify language-minority students and take affirmative actions to rectify the linguistic deficiency so that students can learn. Once English-language learners are identified through assessment, they are regularly assessed to monitor their development of fluent English proficiency (Abedi, 2013). There is significant debate regarding requiring English-language learners to be part of the state accountability system under the No Child Left Behind Act (see Chapter 10). Some educational policy makers argue that English-language learners should take the same accountability test as other learners. Other policy makers claim that achievement testing in English does not assess the students' learning and may contribute to the disproportionate representation of English-language learners in special education. It is clear that in the U.S. educational system, the assessment of English-language learners is a complex and challenging issue. In a comprehensive review of assessment accommodations for English-language learners, Abedi, Hofstetter, and Lord (2004) warned against a one-size-fits-all approach to accommodation.

Instrument Translation, Adaptation, and Modification. Sometimes professionals will consider translating an existing instrument into the language of the nonnative speaker in order to assess the person fairly. A simple translation, however, usually does not suffice, because many issues need to be attended to in the translation. First, simple translations typically do not take into account different cultural considerations. As Bracken and Barona (1991) documented, there are several problems with translating instruments. First, test directions are often not easy to translate because they are often psychotechnical, stilted, and abstract. Second, practitioners rarely translate back to the original after translating the instrument to the second language, which can be a helpful step in determining if the meaning is retained in the translating process. Third, some of the psychological variables assessed in instruments are not universal across cultures. For example, an instrument could assess a person's ability to make decisions independently, which is not a value held by all cultures. Fourth, even for well-developed instruments, there are no universally accepted procedures or standards for evaluating translations. To translate an instrument properly, experts need to evaluate whether the translation is equivalent to the original version of the instrument in terms of content, difficulty level, and reliability. They also need to study the validity of the translation to see if it can be used in the same circumstances as the original instrument. In the instrument's manual, test publishers should describe the process of translating an instrument and provide information on the reliability and validity of the translated instrument. Even if care is taken in translating an instrument, counselors should not assume that clients' cultural experiences are comparable. For example, an item on an interest inventory concerning whether one "likes to go to the opera" may have little relevancy for an adolescent from rural Pakistan.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

These same difficulties apply to attempts to adapt or modify an assessment for a nonnative speaker. Counselors should not attempt to adapt or modify an assessment on their own. Fernandez, Boccaccini, and Noland (2007) recommended the following four-step approach for identifying and selecting translated tests for Spanish-speaking clients. First, identify translated tests by examining assessment publishers' catalogs and websites. Second, identify research studies of translated tests. Fernandez et al., however, noted that the demand for translated instruments has outpaced the ability of test developers to evaluate the psychometric properties of these assessments. Their third step involved confirming that the research applies to the client who will be completing the assessment.

The last step concerns determining the level of research supporting the use of the translated test with the client. Central to fair assessment is the concept of equivalence, and even minor modifications can sometimes alter what is being assessed. If a publisher recommends linguistic modifications, the research supporting these modifications and the modification methods should be described in detail in the manual. If research supports the modifications, practitioners should follow the procedures unless other research indicates that there are new procedures that are more appropriate. Potential problems can be avoided if practitioners thoroughly evaluate the existing research on assessment with nonnative speakers.

Issues Related to Instrument Use and Interpretation. Instruments are available in multiple languages, and with some multilingual individuals, it is difficult to determine which language is most appropriate to use in testing. According to the Individuals with Disabilities Education Act, which was last reauthorized in 2004, individuals should be assessed in the language with which they are most proficient. An exception is allowed when the assessment's purpose is to determine a level of proficiency in a certain language (e.g., English). It is important for counselors to remember that assessing individuals in their most proficient language does not guarantee a comprehensive and fair assessment.

The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) recommends that assessment should be administered in the language that is most relevant and appropriate to the purpose of the assessment. Counselors should consider the client's linguistic and cultural background and the relative language proficiencies of clients who are bilingual or multilingual. The *Standards* state that the examinee should be assessed in language they are most proficient except in the case where the purpose of the testing is to determine the examinee's level of proficiency in a particular language. In some cases, however, the language they are most proficient and the language of instruction vary; in these cases, the examinee should be tested in the language of instruction. The *Standards* clearly state that professional judgment needs to be used in determining the most appropriate procedures for establishing language proficiencies.

L09 Assessment of Individuals with Disabilities

In assessment, individuals with disabilities may be adversely affected by certain assessment procedures. For example, if a person has become quadriplegic, it would be difficult to perform the Block Design portion of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) because of limitations with motor dexterity. The likelihood of a counselor having a client with a disability is high—about 19% of individuals in the United States have some type of disability with more than half of that group reporting that the disability was severe (U.S. Census Bureau, 2012). According to the Individuals with Disabilities Education Act (IDEA), Public Law 101-476, there are 13 separate categories of disabilities under

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 6.1
Thirteen separate categories of disabilities under which children may be eligible for special education and related services

- Autism
- Deafness
- Deaf-blindness
- Hearing impairment
- Mental retardation
- Multiple disabilities
- Orthopedic impairment
- Other health impairment
- Serious emotional disturbance
- Specific learning disability
- Speech or language impairment
- Traumatic brain injury
- Visual impairment

which children may be eligible for special education and related services. These are listed in Table 6.1. As was mentioned in Chapter 5, the Americans with Disabilities Act has legal implications related to the assessment of individuals with disabilities, and many counselors (particularly school counselors) need detailed knowledge about this legislation. As Helms and Ekstrom (2002) reported, counselors often have multiple roles when working with individuals with disabilities—empathic listener, advisor, educator, advocate, and intermediary; hence, knowledge of assessment and the difficult issues associated with "reasonable accommodations" are needed. Counselors may also be involved in documenting a disability or in documenting the accommodations that were implemented during a testing situation (see Vacc & Tippen, 2002).

Finding reasonable accommodations for individuals is important, particularly in relation to achievement tests, because students with disabilities tend not to do as well as their peers without disabilities. Altman, Thurlow, and Vang (2010) found that only 30% of students with disabilities were performing on achievement tests at a level that would be considered proficient. Although the following discussion of specific disabilities addresses certain accommodations, it should be noted that there is not general consensus across states on whether several accommodations, including calculator, read-aloud questions, sign interpreted questions, and spell-checker, can be utilized (Lazarus, Thurlow, Lail, & Christensen, 2009).

An issue in assessment related to testing people with disabilities is the topic of *flagging*, which involves reporting any and all aspects of the test administration or interpretation that deviate from the standard procedure (Pratt & Moreland, 1998) and usually was used in postsecondary assessment like the SAT (Lovett & Lewandowski, 2015). Flagging of test score is a controversial issue, and there are also some legal requirements (AERA, APA, & NCME, 2014). In general, where there is credible evidence that the scores from regular administration and altered tests are comparable, then flagging is *not* recommended. If, on the other hand, if there is credible evidence that the results from the accommodations and the regular administration are not comparable, then consideration of informing score users (e.g., university admissions officers) by potentially flagging the results should be considered. Flagging results should comply with potential laws, and counselors need to be aware of these laws. For more information on flagging and its controversies, readers are directed to Lovett and Lewandowski (2015).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Assessment of Individuals with Visual Impairments

The terms *visually impaired* and *visually handicapped* usually refer to individuals who require more than corrective lenses to function adequately (Bradley-Johnson & Ekstrom, 1998). Many of the assessment tools used in counseling involve reading; yet, for individuals with visual impairments, this type of assessment is problematic. In addition, Bertone, Bettinelli, and Faubert (2007) found even minimal loss of visual acuity (20/40) had some effect on the performance for certain nonverbal tests. With many types of instruments (e.g., personality assessments), the approach to assessing someone with a visual impairment is to have someone read the instrument to the client. Individuals who read assessment material for visually impaired individuals, however, can vary dramatically in terms of their tone, speed, and clarity of speech. These variations in reading can affect the reliability of the instrument and may have an impact on individuals' performance. Thus, as with other assessment practices, counselors need to research the procedures for using a specific assessment instrument with an individual who has a visual disability.

If counselors suspect that a visual impairment might be affecting the assessment process, they should refer their client to an optometrist or ophthalmologist for an initial screening. Some of the signs of a visual impairment are a client who rubs her eyes excessively; shuts or covers one eye, tilts her head, or leans forward; has difficulty in reading or loses her place while reading; blinks more than expected; is irritable when doing close visual work; has difficulty judging distances; or has recurring sties (Sattler & Evans, 2014). Sometimes counselors may need to determine if the assessment conducted with a client who has a visual impairment was comprehensive. In a comprehensive assessment, counselors would expect that the examiner observed the individual functioning in a naturalistic setting and conducted a functional vision evaluation. Kelley, Sanspree, and Davidson (2001) indicated that about 90% of children with visual impairments retain some vision. In assessment, it is important to consider the severity of the visual impairment and the age of onset of the disability. Individuals who lose their vision after the age of five usually retain some visual memories (Bradley-Johnson & Ekstrom, 1998). However, many conceptual factors (e.g., color, facial expressions) are difficult to teach to those who have never had vision. Furthermore, 30–50% of visually impaired children have other impairments (Kirchner, 1983) that need to be considered in the interpretation of the assessment results.

With many well-known instruments, researchers have studied the most effective methods for adapting the instrument for visually impaired individuals to eliminate potential sources of error. For example, for individuals with visual impairment, there are procedures for adapting the Wechsler Intelligence Scale for Children-Fifth Edition (WISC-V) that use more of the subscales for Verbal Comprehension Index. Other instruments include prerecorded tapes that read the instruments in a systematic manner. In addition, many of the personality inventories have Braille or large-print versions. An adaptation of the Stanford-Binet is the Perkins-Binet Tests of Intelligence for the Blind, which is standardized on separate forms for partially sighted and blind children. One of the few instruments developed specifically for visually impaired individuals is the Blind Learning Aptitude Test.

When clients with a visual impairment are being assessed in the areas of personality and career interests, it is important that they answer the questions related to what they enjoy and like to do rather than what they are capable of doing. For achievement and aptitude assessments, Bradley-Johnson and Ekstrom (1998) found that common accommodations for individuals with visual impairments are the following: (a) the provision of special lighting; (b) changes in test scheduling, such as allowing extra time or allowing testing over several days; (c) revision of test directions, such as reading print directions aloud; (d) revision of

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

test format, such as using Braille, large print, or audiotape; and (e) revision of response methods, such as dictating answers or using such low-vision aids as magnification devices, tactile accessories, or computers. It should be noted with Braille instruments, readers usually read it at a slower pace than written language, and time to take the assessments should also be adjusted (Lovett & Lewandowski, 2015).

Assessment of Individuals with Hearing Impairments

Assessing individuals with hearing impairments involves a different set of problems than does assessing individuals with visual impairments. Hearing has a direct link to language acquisition, and verbal content is embedded in many of the methods by which individuals are assessed. In the past, counselors sometimes assessed hearing-impaired children by using instruments that involved reading, but these counselors did not have an understanding of the connection between hearing and reading. It is often difficult to decipher aptitude and ability separately from the verbal disability. In recent years, gains have been made in identifying children with hearing problems earlier and initiating beneficial educational interventions sooner.

In terms of understanding clients with hearing disabilities, counselors need to consider the magnitude of the hearing loss, the age of onset of the problem, the effects of the loss on language acquisition, and whether the loss is affecting adjustment. Hearing disabilities are often measured in terms of the decibel level (sound intensity) needed to recognize the presence of a sound at a given frequency (as measured in hertz; Brauer, Braden, Pollard, & Hardy-Braz, 1998). It is also important to realize that with some individuals, hearing may fluctuate. Furthermore, some within the deaf community object to the terms *hearing impaired* or *hearing loss* because of the pejorative nature or focus on loss. Instead, they prefer the term *deaf*.

Assessment of individuals with hearing disabilities should be conducted in the individual's preferred communication modality or language (e.g., American Sign Language). If a sign language translator is used, the individual's proficiency needs to be considered, and, typically, family members should not serve as translators. American Sign Language is a distinct language and not merely a translation of English into signs; hence, in assessment, the translator will need to make professional judgments that may influence the content of the assessment.

There are also more instruments with norms for individuals with hearing impairments than for those with visual impairments. The Wechsler scales have been used extensively with hearing-impaired children. For example, the WISC-R has established norms for individuals with hearing impairments. In terms of accommodation for a large-scale standardized test, Cawthon (2009) found the most common accommodations were extended time, small group/individual administration, test directions interpreted, test items read aloud, and test items interpreted. It should also be noted that impairments of the vestibular system can affect balance, equilibrium, and other motor skills. In terms of assessment in the affective domain, counselors need to ensure that clients with hearing impairments comprehend the questions. Sometimes, particularly with children, clients may give the impression that they understand what is being discussed when, in fact, they are confused. There are also some indications that individuals with hearing impairments have a tendency to respond in concrete terms to projective techniques or to more abstract counseling questions.

Assessment of Individuals with Motor Disabilities

Imagine for a minute that even though you understand what is being asked of you on a test, but you are unable to make your body physically respond and complete the required task. Individuals with motor disabilities often face this exact situation because they may

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

immediately realize how to put the puzzle pieces together but are unable to do so quickly. The methods used to assess clients with motor impairment, as with those used for all clients, vary depending on the information needed. Sometimes the intent of the testing is to assess where a client's strengths and limitations are to determine possible careers and to identify areas where accommodations can be made.

Under certain circumstances, examiners will perform assessments with clients with motor disabilities twice; once using the specified time limits and a second time without time limits to determine if clients can perform the tasks. A well-written psychological report will identify any departures from the standardized procedures and will discuss how the violations might have influenced the results.

With assessment of individuals in rehabilitation settings, vocational evaluation can involve three approaches to assessment: psychological testing, use of work sample or related work activities, and actual on-the-job activities. Because individuals with physical disabilities often tire more easily, accommodation often involves testing in time-limited intervals and over several days. Many times computers are used so that individuals with motor disabilities can complete instruments with little outside assistance. Examples of computer technologies used are voice inputs, simplified keyboards, pneumatic controls, joysticks, head pointers, and Braille keyboards. These devices can eliminate an intermediary who reads or responds to the test items for the examinee. For more information on assessment of individuals with physical disabilities, readers are directed to Frank, Rosenthal, and Caplan (2009).

There are a number of methods for assessing intelligence or general ability of individuals with motor disabilities. One commonly used method involves picture vocabulary tests, in which the individual simply points to the correct answer. The pictures eliminate the problems associated with reading, and the pointing is not timed and eliminates many of the problems associated with motor coordination difficulties. The *Peabody Picture Vocabulary Test-Fourth Edition* (PPVT-IV; Dunn & Dunn, 2007) is the most commonly used of these instruments. The PPVT-IV takes about 10 to 15 minutes to administer. The examinee is presented with a plate with four pictures. The examiner says a stimulus word, and the examinee responds by pointing to the appropriate picture. PPVT-IV scores are similar to other measures of intelligence, with a mean of 100 and a standard deviation of 15. Another instrument, the Expressive Vocabulary Test, measures expressive vocabulary and word retrieval and is co-normed with the PPVT-IV to allow for comparisons of receptive and expressive vocabulary.

Assessment of Individuals with Cognitive Impairments

The assessment of individuals with intellectual disabilities (formerly known as mental retardation) falls within the area of assessment of cognitive impairments. Counselors usually do not perform testing related to the determination of intellectual disabilities, but it is important for them to have an understanding of this area. The American Association on Intellectual and Developmental Disabilities (AAIDD) defines an intellectual disability as a disability characterized by significant limitations both in intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills in which the disability originates before the age of 18. The AAIDD definition of intellectual disability is similar to the *Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition-Text's* (DSM-5; American Psychiatric Association, 2013) criteria for intellectual disability (intellectual developmental disorder; see Chapter 15). Using the *DSM-5* criterion, intellectual disability is characterized by deficits in intellectual functions confirmed by clinical assessment and individualized, standardized intelligence testing. The intelligence deficits must be accompanied by concurrent deficits or impairments in adaptive functioning. In the *DSM-5*, the onset of both the intellectual and adaptive deficits

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

must occur during the developmental period. Therefore, both the AAIDD definition of intellectual disability and the *DSM-5* criteria focus on limitations in both intellectual and adaptive functioning.

The assessment of an intellectual disability typically involves an individual test of intelligence (e.g., the WISC-V or the Stanford-Binet) and a measure of *adaptive* behavior. Psychologists often used the *Vineland Adaptive Behavior Scales-Second Edition* (Vineland-II) to measure adaptive behavior. There are four versions of the recently published Vineland-II that can be used in combination or alone: the Survey Interview Form, Parent/Caregiver Rating Form, the Teacher Rating Form, and the Extended Interview Form. The major domains assessed in the Vineland-II are Communication, Daily Living Skills, Socialization, and Motor Skills. There is also an option to score an Adaptive Behavior Composite. The AAIDD has published the *Supports Intensity Scale*, which is designed to assess the practical support requirements of an individual with intellectual disability. This assessment is somewhat unique in that it focuses on what the individual may need to best function rather than on identifying deficits. AAIDD is also scheduled to release the *Diagnostic Adaptive Behavior Scale* (DABS) in 2015, which is purported to provide a comprehensive assessment of adaptive behavior.

The AAIDD has developed guidelines for professional conduct for practitioners working with individuals with intellectual and developmental disabilities, which are:

1. The practitioner objectively solicits, honors, and respects the unique needs, values, and choices of the persons being served.
2. The practitioner communicates fully and honestly in the performance of his or her responsibilities and provides sufficient information to enable individuals being supported and others to make their own informed decisions to the best of their ability.
3. The practitioner protects the dignity, privacy, and confidentiality of individuals being supported and makes full disclosure about any limitations on his or her ability to guarantee full confidentiality.
4. The practitioner is alert to situations that may cause a conflict of interest or have the appearance of a conflict. When a real or potential conflict of interest arises the practitioner not only acts in the best interest of individuals being supported but provides full disclosure.
5. The practitioner seeks to prevent and promptly responds to signs of abuse and/or exploitation, and does not engage in sexual, physical, or mental abuse.
6. The practitioner assumes responsibility and accountability for personal competence in practice based on the professional standards of his or her respective field, continually striving to increase professional knowledge and skills and to apply them in practice.
7. The practitioner exercises professional judgment within the limits of his or her qualifications and collaborates with others, seeks counsel, or makes referrals as appropriate.
8. The practitioner fulfills commitments in good faith and in a timely manner.
9. The practitioner conducts his or her practice with honesty, integrity, and fairness.
10. The practitioner provides services in a manner that is sensitive to cultural differences and does not discriminate against individuals on the basis of race, ethnicity, creed, religion, sex, age, sexual orientation, national origin, or mental or physical disability.

Another major area of assessment related to cognitive impairments is **neuropsychological assessment**, which concerns assessing the behavioral effects of possible brain injury or damage. Brain damage can result from injury or disease, or it can be organic. Insults to the nervous system are often associated with cognitive dysfunctions, such as memory difficulties, but they can also have an effect on personality and temperament. Once again,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

although counselors will not be performing neuropsychological testing, they do need an overview of the typical process to know when to refer clients for neuropsychological assessment. Clients who display abrupt mental or behavioral changes should be referred for a neuropsychological assessment. An evaluation is also probably warranted if the changes are more subtle, such as individuals exhibiting mild memory changes, borderline EEG (electroencephalograph) results, attention lapses, speech disturbances, and motor dysfunction. It is also wise for a client to be evaluated if there is any suspected trauma to the brain resulting from illness or injury. For instance, clients should be referred for a neurological screening if an injury resulted in loss of consciousness.

Neuropsychologists often administer a battery of tests to clients because no single test can detect the diversity of brain dysfunctions. These batteries were designed to systematically assess a broad spectrum of neurological functioning. Two of the most commonly used assessment batteries for neuropsychological testing are the *Halstead-Reitan Neuropsychological Test Battery* and the *Luria-Nebraska Neuropsychological Battery*. The Halstead-Reitan takes about 6 to 8 hours to administer and typically involves administering either the WISC-V or the WAIS-IV. The Luria-Nebraska takes about 2.5 hours to administer and is designed to assess a broad range of neurological impairments. These specifically designed neuropsychological batteries are not the only methods used in this area. Neuropsychologists will also often use individual measures of intelligence or specific subscales (e.g., the Digit Span of the WAIS-IV) as part of the assessment process. Another commonly used instrument is the *Bender Visual Motor Gestalt Test*, usually referred to as the Bender-Gestalt Test. The Bender-Gestalt Test consists of nine cards, each of which contains an abstract design that the examinee is asked to copy. The Bender-Gestalt may reveal visual-motor difficulties indicative of neurological impairments. There are different methods of administering the Bender-Gestalt, many of which require that the examiner show the examinee each card for a specific period of time so that information on memory deficits can be obtained. In addition to neurological assessment, the Bender-Gestalt is also frequently used to assess possible learning disabilities.

Recent medical advances have also contributed to the assessment of brain injury and damage. Techniques such as magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI) provide better insight into the brain's functioning than ever before. These medical techniques, however, cannot provide the entire picture, and often neurologists and neuropsychologists work together, both in diagnosing the impairments and in determining appropriate treatment. Sometimes the treatment involves counseling, either for the individual with the neurological impairment or for the family coping with the disability.

Assessment of Older Individuals

Although not a disability, I have chosen to insert a section on assessment of older individuals because there is a significant likelihood that a counselor will work with an individual over the age of 65 given that this is a growing segment of the population. The number of people today aged 60 and over has doubled since 1980 (World Health Organization, 2012). The American Psychological Association (2013) developed the *Guidelines for Psychological Practice with Older Adults* because of these demographic changes, increased knowledge related to the aging process, and refined assessments with older adults. Within these guidelines for psychologists is a section devoted specifically to assessment. These guidelines also address the need for specific training related to conducting and interpreting cognitive screening and functional ability evaluations. Although most counselors will not be conducting cognitive screenings and functional

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

ability evaluations, they may be working on multidisciplinary teams in which such information is useful.

In assessing older adults, clinicians need in-depth knowledge of the normal aging process in order to identify deviations from these normal processes. Counselors may be involved in the initial screening of older adults for problems with cognitive processing or for cognitive deficits. As mentioned in Chapter 8, the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975) is used in a mental status examination. The MMSE is a brief 30-point questionnaire that screens for cognitive functioning in which the client is first asked a series of questions (e.g., "What is the month?" "What is the year?" "What is the date?" "What season is it?"). Other items involve seeing if they can respond to a command (e.g., "Take the paper in your hand, fold it in half, and put it on the floor"), and read and respond to a phrase (e.g., "Close your eyes"). The exam ends with the examinees copying two intersecting pentagons. Lopez, Charter, Mostafavi, Nibut, and Smith (2005) found that many of the items on the MMSE were too easy. Readers should be aware that there is now a second edition of the MMSE.

With older adults, some psychological assessments will focus on memory issues because of the prevalence of memory-related illnesses (e.g., Alzheimer's, dementia). This is an area of assessment that requires specialized knowledge and training. It also should be noted that clinically significant problems are not restricted to older clients and that memory problems can occur at any point in the lifespan (e.g., after a head injury). A commonly used assessment is the Wechsler *Memory Scale-Fourth Edition* (WMS-IV; Wechsler, 2009). The WMS-IV is an individually administered instrument designed to measure visual, auditory, and working memory. The WMS is neurologists' third most frequently used clinical assessment (Camara, Nathan, & Puente, 2000). It contains two batteries: the Adult battery for individuals aged 16–69, and a shorter Older Adult battery for individuals aged 65–90. The WSM-IV is co-normed with the WAIS-IV and can be given concurrently with the WAIS-IV for a more comprehensive assessment. The Adult battery has seven subtests yielding five index scores: Auditory Memory, Visual Memory, Visual Working Memory, Immediate Memory, and Delayed Memory. The Older Adult battery was purposely designed to be shorter and has five subtests that yield four index scores. Chittooran (2010) evaluated the WSM-IV as being research and theory based, statistically sound, and having excellent clinical applications. She also evaluated the reliability evidence as indicating notable internal consistency, excellent interscorer reliability, and good test-retest stability.

Certain psychologists may also specialize in assessing functional impairment or behavioral disturbances with older adults (APA, 2013). Hood and Johnson (2007) contended that assessment of functional impairment should include both assessment of cognitive functioning and health behaviors. In the health care field, the term *activities of daily living* (ADLs) refers to the basic tasks of everyday life, such as eating, bathing, dressing, and toileting, and are often considered in assessment of elderly patients. ADLs are predictive of admission to a nursing home (Branch & Jette, 1982); use of paid home care (Garber, 1989); and use of hospital services (Branch, Jette, & Evashwick, 1981). Not all surveys or checklists use the same list of ADLs, but most surveys include self-feeding, bathing, dressing, and grooming.

Some mental health issues may have unique presentations in older adults. As an example, depression may be difficult to distinguish from coexisting cognitive impairments and may not be expressed as overt sadness. Moreover, age is also associated with loss of friends and family, so it may be difficult for clinicians to separate issues related to the normal grieving process from depression. The Geriatric Depression Scale (GDS; Yesavage et al., 1983) is a commonly used 30-item scale that was developed expressly for assessing depression with older individuals. A shortened, 15-item version was also developed

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(Lesher & Berryhill, 1994) to provide a quick screening for depression with this population. Anxiety and anxiety-related disorders may also be a relevant clinical issue for some older adults. Furthermore, clinicians should also be sensitive to possible substance-use issues with older adults as problems in this area can exist for many older individuals (Olin & Blow, 2000). Gaining an understanding of the clinical problems may involve assessment of behavioral patterns (e.g., dependency, aggression) and contextual factors (e.g., levels of social support or family interactional patterns) (APA, 2013).

Comprehensive assessment of older clients can be complex and may often involve a multidisciplinary approach. When assessing older individuals, counselors should be sensitive to issues of fatigue and the possible effects of medications (Hood & Johnson, 2007). Furthermore, counselors need a thorough understanding of the psychological, developmental, and physiological aspects of aging.

Summary

Counselors can expect to work with clients from various backgrounds; therefore, they need to be skilled in cross-cultural or multicultural assessment. When selecting an instrument, counselors need to consider the client's background and select an instrument appropriate for that client. Counselors also work in locations (e.g., schools) where they do not choose the assessment instruments (e.g., achievement tests) but are still responsible for administering and interpreting the results in a fair and unbiased manner. One resource that counselor may want to consider using is the *Standards for Multicultural Assessment*, which includes a broad definition of multicultural assessment and it addresses issues of social advocacy in assessment.

In all assessment situations, counselors need to evaluate the assessment instruments for possible bias. One area counselors should consider is whether the instrument has been evaluated for possible content bias. Content bias concerns whether the content of the instrument is more familiar or appropriate to one group as compared with another group or groups. Another area counselors should examine is the instrument's internal structure for different ethnic groups. Instruments can differ in the degree to which they are reliable for different groups, and the factor structure may vary depending on the ethnic group. The analysis of possible bias should also include studies related to the relationship between the instrument and the criterion for different groups. It is important to examine differences among the validity coefficients and possible slope bias. Another area to examine is whether there are differences in the intercepts for different ethnic groups.

Some of the instruments counselors use may not be biased but may still produce results that have a disparate or negative impact for some racial or ethnic groups. There is debate about whether differences in ability, achievement, and aptitude testing are primarily genetic, social, or cultural. Presently, it is impossible to develop a culture-free instrument, but many professionals are committed to developing instruments that attempt to be more culturally fair.

Counselors also need to consider clients with disabilities. In assessing clients with disabilities, counselors need to review the research on the most appropriate methods for adapting the assessment for these individuals. In many settings, a counselor may be on a multidisciplinary treatment team that will be involved with developing treatments for individuals with disabilities. The client may have been assessed by a vocational rehabilitation specialist or a psychologist, and the counselor will need to understand the psychological assessment report. A well-written psychological report will identify any departures from the standardized procedures and will discuss how the violations may influence the results.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The results from any significant departure from standard procedures should be viewed with concern. Assessment with a special population can provide useful information if it is carefully performed and conservatively interpreted.

Counselors should also be knowledgeable about assessment issues related to older individuals as this is a growing segment of the population. For many types of impairments, such as cognitive or memory impairments, counselors will not be the ones conducting the assessments, but they may be part of a multidisciplinary team that treats older adults. Given the complexities of aging, counselors working with geriatric clients need an understanding of the normal aspects of aging and specialized training in respect to mental health issues with this population.

In conclusion, assessment with individuals from diverse backgrounds and cultures is a complex topic, and it is anticipated that this topic will continue to be debated in the next 10 or more years. The goal of any assessment is to provide pertinent information that is meaningful, fair, and accurate. This is often a challenging goal because the constructs counselors are attempting to assess are typically multifaceted. There are, however, many dedicated professionals who are attempting to address topics related to assessment with diverse individuals, and it is anticipated that in the coming decades these professionals will make strides in leveling the playing field related to many types of assessments.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 7

Selecting, Administering, Scoring, and Communicating Assessment Results

Learning Objectives

- LO1 List the steps in selecting an assessment
- LO2 Understand the importance of knowing the administrative procedures of an assessment
- LO3 Describe the terms *performance assessment* and *authentic assessment*
- LO4 Explain effective methods for communicating results to clients
- LO5 Discuss effective methods for communicating results to parents
- LO6 Describe the common areas in a psychological report
- LO7 List the attributes of a well-written psychological report

Before moving on to next section of this book, which addresses specific assessment areas, it is important to address the specific topics of selection, administration, scoring, and communicating assessment results. Particularly in terms of the use of formal instruments, problems related to assessment can often be traced to clinicians' confusion or misjudgments in one of these areas. This discussion of selecting, administering, scoring, and communicating results may seem to reinforce the idea that assessment is a separate and distinct process for counseling. My philosophy, however, is that counselors are continually assessing clients throughout the counseling process and that clinicians need to be mindful of the appropriate use and interpretation of both informal and formal assessment techniques.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

LO1 Selection of Assessment Instruments or Strategies

A practitioner who selects and uses a faulty assessment tool in counseling could adversely affect a client's life. A glaring example would be to mishandle an assessment of suicidal ideation. Therefore, the selection of an assessment tool or strategy is an important clinical decision. In some counseling situations, the selection of assessment strategies is neglected due to heavy caseloads or comfort with existing methods. Yet, assessment is the foundation of the counseling process because it provides much of the information used to determine the therapeutic goals and interventions. Therefore, the selection of assessment methods needs to involve careful consideration and evaluation of strategies and approaches. The following steps may be helpful in selecting an assessment instrument or strategy: (1) determine what information is needed, (2) analyze strategies for obtaining information, (3) search assessment resources, (4) evaluate assessment strategies, and (5) select an assessment instrument or strategy.

Determine What Information Is Needed

The first step in the process of selecting an assessment strategy is to determine what information would be useful. An assessment tool is helpful only if it provides useful information. Therefore, before counselors can determine what assessment strategies to further consider and evaluate, they first need to determine what information is truly needed. Doing so can involve identifying the information needed for a specific client or determining what general information clinicians in an organization may need about their clients. Sometimes counselors skip this first step and implement an assessment approach without considering whether the process will yield useful information. As reflected in the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practice, 2004), the first responsibility for test users is to "Define the purpose for testing, the content and skills to be tested, and the intended test takers. Select and use the most appropriate test based on a thorough review of available information" (p. 5). According to Anastasi and Urbina (1997), the use of formal assessment is most productive when it answers a client's question and is seamlessly integrated into the counseling process. Sometimes when counselors are considering assessment strategies and client information that would be useful; they do not consider information that may already be accessible. For example, school counselors often have access to substantial information in students' educational records, and mental health counselors frequently have clients complete some forms before an initial session. Hence, sometimes information can be accessed in advance and the assessment process can focus on strategies for eliciting particularly pertinent client information. It should also be stressed that gathering assessment information is not applicable just at the beginning of counseling; counselors often need to determine assessment strategies at various points throughout the counseling process.

Analyze Strategies for Obtaining Information

Once clinicians have identified the pertinent client information, the second step is to determine the best strategies for obtaining that information. A primary decision for a counselor is whether information can best be gathered through *formal* or *informal* techniques. Frequently, both formal and informal assessment tools provide an optimum situation, because the strengths of one method complement the other, and together they can provide an in-depth analysis of clients. Counselors also need to consider which assessment method would be most suited to a client or clients (e.g., paper-and-pencil self-report, behavioral checklist, computer assessment). Often counselors will become comfortable with certain assessment strategies (e.g., interviewing) and not consider other strategies

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

that may correspond to client characteristics. Analyzing client attributes before selecting an assessment strategy may increase both the quality and quantity of the information gathered. Furthermore, counselors also need to consider their own professional limitations and the types of instruments they can ethically administer and interpret.

Search Assessment Resources

Once the pertinent information and the most effective strategies for gathering that information are identified, the next step is to peruse existing instruments or assessment strategies. It is impossible for any practitioner to be familiar with the thousands of published and unpublished assessment tools in the counseling realm. In addition, it is not always easy to identify possible instruments to use. Because no one source describes every formal and informal assessment tool that has ever been developed or used, practitioners often need to use a variety of sources and technologies to develop a list of possible tools. Counselors should regularly examine assessment resources because new instruments are continually being developed that may provide better information. In addition, many instruments are continually revised and a counselor would not want to use an older version of an instrument when a more recent and improved version is available. Although it may be tempting to use the Internet and some of the more popular search engines (e.g., Google) to identify assessment resources, this approach has some problems. I recently used Google to identify career interest inventories and it identified 8,010,000 websites. Even though I do find the examination of new interest inventories quite fascinating, even I am not going to visit and thoroughly evaluate all 8,010,000 sites! Furthermore, based on the information provided in some of the websites, it is somewhat difficult to differentiate between legitimate, psychometrically sound interest inventories and ones with little evidence regarding the instrument's psychometric qualities. Therefore, using some of the resources discussed in the next section can be highly useful.

Assessment Search Resources. A number of resources can assist counselors in searching for different assessment instruments. One of the most important assessment resources is the **Mental Measurements Yearbooks** (MMY). This series of yearbooks contains *critiques* of many of the commercially available psychological, educational, and career instruments.

Oscar K. Buros published the first yearbook in 1938 and continued to spearhead the publication of subsequent yearbooks until his death in 1978. Each yearbook extends information in the previous yearbooks by reviewing instruments published or revised during a specific time period. The *Nineteenth Mental Measurements Yearbook* is the most current edition of the yearbooks. The Buros Institute of Mental Measure can be accessed online at http://www.buros.org. Most academic libraries have book forms of the *Mental Measurement Yearbooks* and test reviews can be retrieved either online or through a library. By using the Buros Institute home page, clinicians can use "Test Reviews Online" in which counselors can search a large database of assessments. For example, typing in the term *anxiety* recently brought up a lengthy list of assessments that measure anxiety. By clicking on an instrument's name, a counselor can then get specific information on the instruments, including title, author, purpose, acronym, publisher, publisher's address, and in which edition of the *Mental Measurements Yearbook* the assessment was reviewed. It is also possible to order the review (for a fee) directly from the website. These reviews and critiques are written by authorities in the assessment field and provide an overview of each instrument's strengths and limitations. Another resource published by the Buros Institute is *Tests in Print*, which is useful in identifying tests by content area. The current version, *Tests in Print VIII* (Murphy, Geisinger, Carlson, & Spies, 2011), summarizes information on approximately 3,000 testing instruments. The instruments listed in *Tests in Print* are also cross-referenced with the Mental Measurements Yearbooks.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Another good resource for identifying possible instruments is Educational Testing Service (ETS) Test Link, which also can be accessed online. Test Link includes the Test Collection at ETS, which includes information on more than 25,000 tests and other measurement devices. Many of these are author-developed instruments that have been used in research; however, some are applicable for practice also. Many of these instruments can be downloaded for a fee from the ETS Test Link website.

Another online database that focuses on unpublished instruments that have been developed by researchers is PsycTESTS. PsycTESTS is a product of the American Psychological Association and a subscription may be available for students through their university library resources. Similar to other databases, you can enter an assessment term (e.g., depression) and PsycTESTS will provide lists of assessment related to depression. Then the user can select one or more of the instruments and most records will include a description of the test and links to peer-reviewed literature, technical reports, or dissertations that are related to the instrument.

Evaluate Assessment Strategies

After identifying possible assessment strategies, the next step is to evaluate those strategies in order to select the most effective methods or instruments. One of the first steps in determining if an instrument is appropriate to use is to ascertain if the instrument's purpose corresponds to the counselor's needs. If an instrument does not measure the behavior or construct of interest, then there is no need to further investigate the psychometric characteristics of that instrument. As mentioned earlier, an instrument is worthwhile only if it provides pertinent information that is needed in the counseling process. Sometimes, determining an instrument's purpose involves sifting through the information provided in the manual. Occasionally, the purpose stated in the manual may be a little different from what the name of the instrument implies. Furthermore, it is important to determine if the design of the instrument corresponds to the intended purpose. Some instruments fail to achieve their intended purpose because of problems in construction. An instrument's manual should provide a description of the instrument's purpose as well as a detailed explanation of how the instrument measures the behaviors related to that purpose.

Instrument Development. Counselors need to closely scrutinize the procedures that instrument developers used in constructing the instrument. Most instruments have a theoretical or empirical base that influences the overall development. If detailed information about the development of an instrument is not provided, counselors should be concerned, and its psychometric properties should be carefully evaluated. For many instruments, the development of the actual items plays a central role in instrument construction. For example, if an instrument is based on item response theory, the development and analysis of each item is critical to the interpretation of results. With any assessment, evidence of careful construction of items, as well as of detailed item analysis, adds credence to the instrument's value. Counselors should also read the items and consider if they are appropriate for the client(s). Furthermore, if counselors will use the instrument with clients from different ethnic groups, then they should examine whether *differential item functioning* has been conducted to control for group differences (see Chapter 6).

Appropriate Selection of Norming Group or Criterion. Depending on the type of instrument, either the norming group in a norm-referenced instrument or the criterion in a criterion-referenced instrument will need to be evaluated. In a norm-referenced instrument, an analysis of the makeup of the norming group is critical. First, counselors must determine if the norming group is appropriate for their clients. For example, if a counselor is working with adults who did not complete high school, then an instrument that used college students

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

as the norming sample for adults would probably not be appropriate. Counselors should also evaluate the suitability of the norming sample in terms of age, gender, racial or ethnic makeup, socioeconomic status, educational level, and geographic representation. A norming group's racial or ethnic makeup is of particular importance and should be carefully considered. The size of the norming sample is another consideration in evaluating a norm-referenced instrument. Although there are no clear guidelines for determining the appropriate size of a norming group, large samples drawn in a systematic manner tend to garner more confidence.

The evaluation of the criterion in a criterion-referenced instrument involves analyzing the procedures the developers used to determine the criterion. As you will recall, criterion-referenced instruments measure whether an individual has attained a certain level or standard (the criterion). With a criterion-referenced instrument, the evaluation of the criterion is the most important step. In response to legislation and other influences in our society, there is currently substantial interest in setting standards or determining the criteria for passing achievement tests. Publishers of some achievement tests will include information pertinent to setting a criterion for passing that may include standards from professional organizations, reviews by leading researchers, and studies of curriculum. The manual should supply sufficient documentation to convince the reader that the determination of the criterion was based on suitable information and careful analysis.

Reliability. Evaluating an instrument's reliability evidence is an important step in determining whether to use it. When I am teaching an assessment course, counseling students sometimes ask, "How high should a reliability coefficient be to indicate adequate reliability?" My answer is, "It depends." There are no clear guidelines in determining what constitutes acceptable reliability. The type of instrument affects how the reliability coefficients are evaluated. In the next section of this book, the reliability coefficients of some commonly used instruments are provided in order to furnish a reference. Reliability can be estimated in different ways, and in some instruments' manuals, counselors will frequently find a range of coefficients that were calculated using different methods. In comparing instruments, counselors should take into consideration the methods used for determining the coefficients because some methods of determining reliability tend to have lower estimates than others. Another trend to consider when evaluating an instrument's reliability is that reliability coefficients tend to be lower with younger participants. Therefore, you can expect lower coefficients with preschoolers than with adults. In a well-developed instrument, a counselor can expect to find standard error of measurement as a part of reliability information.

Validity. Evaluating the validity of an instrument involves studying the validation evidence. In examining the validation information, counselors should not expect a definitive conclusion, such as the instrument either has validity or does not have validity. Rather, the evaluation of validity involves determining the degree to which there is support for particular *uses* of the instrument. The analysis of validity concerns whether there is a preponderance of evidence to support the use of an instrument in a specific manner or context. The manual should include ample evidence that the instrument is valid for the desired purpose. If there is not ample evidence, the instrument should be used very cautiously or not at all.

Fairness. As was discussed in Chapter 6, fairness in assessment is an important topic and one that a counselor should consider in selecting an instrument. An instrument's authors should provide evidence that the instrument was evaluated for fairness. Chapter 6 provides an overview of how instruments are typically examined to detect possible bias, and counselors should be familiar with these strategies. Because some instruments may be inappropriate for certain populations, the manuals should address those limitations. If a manual recommends that an instrument not be used with certain clients, then alternate assessment strategies should be explored.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Interpretation and Scoring Materials. If an instrument has adequate psychometric qualities, counselors will want to examine the materials that were designed to assist in the interpretation of the results. The manual or additional resources should provide the necessary information for interpreting the results. Many instruments have a variety of scoring options and interpretative reports.

In examining the interpretative materials, counselors also need to consider how they are going to use the information in counseling and what would be the most effective way to share the information with the client. Some instruments have interpretative handouts, which can be given to clients to assist in the counseling process. These handouts should never replace the counselor explaining the results, but these types of resources can sometimes be helpful.

Another consideration when choosing an instrument is how it is scored. Today, many instruments are scored by computer or over the Internet, which may require access to appropriate computers. An organization's financial resources can often affect which instruments and scoring services can be considered for use.

User Qualifications. In selecting an instrument, practitioners should also consider their own competencies and training. Counselors can select instruments only from the pool of instruments that they can ethically use. Moreover, counselors must become competent in using a specific instrument before giving it to a client. Instruments vary in the amount of training and experience that is necessary. Some instruments can be used with knowledge gained from studying the manual, whereas other instruments require supervised instruction on the appropriate use of those devices.

The *ACA Code of Ethics* (American Counseling Association, 2014) indicates that counselors need to recognize their level of competence and perform only those testing and assessment services for which they have been trained (see Appendix A). These standards also indicate that the ultimate responsibility for appropriate use of an instrument rests with the practitioner. Therefore, if an organization wants counselors to use instruments for which they are not properly trained, the counselors have an ethical responsibility to refrain from using those assessments. There can be legal ramifications (e.g., losing a professional license or being sued for malpractice) for practitioners who use instruments that they are not qualified to use.

One of the arguments related to restricting testing and assessment only to psychologists is that unqualified users are using psychological instruments. Qualified and competent use requires more than having the necessary training—the practitioner must also be knowledgeable about the specific instrument being used. The discussion of instruments in the following chapters of this book does not provide enough information to serve as training for using an instrument. Sufficient knowledge of an instrument is gained only through the thorough examination of the manual and the accompanying materials. In some cases, a practitioner also must have supervised training on the use of a particular assessment before being allowed to use it. The counselor must know the limitations of an instrument and for what purposes the instrument is valid.

Practical Issues. When selecting an assessment, there are often a number of practical issues to keep in mind. In many settings, the cost of the instrument plays a role in the decision process. Certainly, a counselor should consider whether the instrument provides information that is sufficiently useful to warrant the costs. Some instruments are quite expensive; however, if used appropriately, they can be enormously helpful to clients. Another consideration is time and whether the time investment is worth it. Some clients may be reluctant to spend hours taking numerous assessment tools, or some parents may object if instructional time is used for assessment. If, however, the clients understand the benefits of the assessment, they may be less concerned about the amount of time involved.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In conclusion, the selection of an assessment instrument is an important task. Assessment tools have been used inappropriately in the past, and sometimes this was because professionals did not devote sufficient time to selecting an appropriate tool for a specific client. A selection decision should not be made without thoroughly examining an instrument and its manual. Figure 7.1 provides some questions to consider when evaluating assessment instruments in counseling.

FIGURE 7.1
Example of a form to evaluate assessment instruments

1. Is the purpose of the instrument clearly stated in the manual?	Yes _____	No _____
2. Does the purpose of the instrument meet my clinical needs?	Yes _____	No _____
3. Do I have the professional skill and background necessary to competently administer and interpret this instrument?	Yes _____	No _____
4. Given the logistics of my organization, can the instrument be administered in an appropriate manner?	Yes _____	No _____
5. Is there evidence that the authors developed clear goals and objectives (i.e., instrument specifications) for the instrument before the instrument was developed?	Yes _____	No _____
6. Was care taken in constructing and evaluating all of the items?	Yes _____	No _____
7. Did they perform adequate item analysis of all items?	Yes _____	No _____
8. Did they evaluate the items fairness or possible bias?	Yes _____	No _____
9. For norm-referenced instruments, does the norming group seem appropriate for my client(s)?	Yes _____	No _____
10. For norm-referenced instruments, does the size of the norming group appear sufficient?	Yes _____	No _____
11. For norm-referenced instruments, does the norm group appear to represent the pertinent population in terms of the appropriate characteristics?	Yes _____	No _____
12. For criterion-referenced instruments, is the criterion clearly stated?	Yes _____	No _____
13. For criterion-referenced instruments, was the criterion selected appropriately?	Yes _____	No _____
14. For criterion-referenced instruments, is there compelling evidence that the level of expected performance is appropriate?	Yes _____	No _____
15. Has the reliability of the instrument been explored sufficiently?	Yes _____	No _____
16. Is there reliability information for different groups and are the coefficients acceptable for these different groups?	Yes _____	No _____
17. Does the instrument appear to have acceptable reliability for my uses?	Yes _____	No _____
18. Does the manual provide the standard error of measurements for all the appropriate scales and subscales?	Yes _____	No _____
19. Does the manual provide validity information?	Yes _____	No _____
20. Is there sufficient validation evidence for me to feel comfortable in using this instrument for the specific purpose needed?	Yes _____	No _____
21. Does the validity evidence indicate I should be careful in using this instrument with certain clients or in certain situations?	Yes _____	No _____
22. Can the instrument be scored in a manner that fits the situation?	Yes _____	No _____
23. Is there sufficient information and resources for me to feel I can adequately interpret the results of this instrument to each client?	Yes _____	No _____
24. Are there appropriate materials to help the client understand the results?	Yes _____	No _____
25. Can the organization afford to use this instrument?	Yes _____	No _____
26. Do the benefits of this assessment outweigh the costs and time commitments?	Yes _____	No _____

Digital Download **Download at CengageBrain.com**

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L02 Administering Assessment Instruments

Many counselors, particularly those in school settings, are responsible for administering certain assessment instruments. This is an important task because assessment results can be invalidated by careless administration. As the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) indicate, "the usefulness and interpretability of test scores *require* that a test be administered according to the developer's instructions" (p. 111). Thus, the major focus of this section is on *knowing the administration materials and being prepared!*

It is important for professionals to have read the administration materials before the time of administering the instrument. With many instruments, pretesting procedures must be completed in a required manner. For example, the administration procedures might require that the room be set up in a specific way to minimize the opportunity for examinees to see others' answer sheets. Other procedures might include specific registration procedures in which the examinee must provide two forms of identification to sit for the examination. Thus, preparedness for the administration of an instrument should start well before the actual administration.

Administrators also need to be aware if there are precise instructions for administering the assessment. Some instruments have detailed instructions on how the instrument is administered, often including a specific script that is read to the examinee(s). In addition, they need to be familiar with testing materials and equipment, and sometimes there are even specific rules about whether the test booklets or the answer sheets are disseminated first. For some school counselors, the administration of state-wide achievement tests has become quite specific, and there are strict rules related to overseeing the testing process. Adherence to a time schedule is often part of the administration responsibilities. For example, a careless mistake in timing on the SAT could significantly affect students' scholarship awards and admission into certain colleges. The *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) indicate that any modifications to standardized test administration procedures need to be documented. Some standardized instruments will provide a form on which the test administrator can report any problems or unusual circumstances that occurred during the administration.

An individual administering an assessment instrument also needs to know the boundaries of what is acceptable. For example, instruments vary in terms of what is considered an appropriate relationship and the degree of rapport the administrator is to have with the examinee(s). Some instruments require a certain degree of reserve, whereas other instruments encourage the administrator to be quite warm and encouraging. Examples of knowledge about what is acceptable include knowing what to do if people have to go to the restroom, procedures if the administrator suspects cheating, or testing protocol if an examinee comes in 30 minutes late. An example of a recent administration problem I witnessed occurred during a scholastic aptitude test, when some administrators allowed students to use calculators and other test administrators did not. The importance of knowing the administrative procedures cannot be emphasized enough. Following the procedures provided in Figure 7.2 may assist counselors who are administering an assessment for the first time. In summary, the key to effectively administering any assessment is knowledge of the procedures and preparing for possible problems.

Scoring

Many instruments are scored by hand, by computer, or over the Internet. Some instruments are self-scored, meaning the client is able to score the instruments, typically by adding columns together or counting responses. An example of an instrument that can be self-scored is the *Self-Directed Search* (Holland & Messer, 2013) on which clients simply

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 7.2
Sample checklist for instrument or test administration

Pretesting Procedures

_____ Ensure that all testing materials are kept in a secure place
_____ Read instrument manual to familiarize yourself with all administrative procedures
_____ Check possible testing facilities for size, seating, lighting, and ventilation
_____ Schedule testing facilities
_____ Check to ensure there are no conflicts with arranged testing time(s) and facilities
_____ Send out any pretesting materials that need to be sent
_____ Review testing materials to ensure that all needed materials are available
_____ Arrange for any posting of signs to direct examinees to testing location
_____ Thoroughly study exact procedures for administering the instrument
_____ Read and practice the directions for administering the test
_____ Check stopwatch or other device if time limits are involved
_____ Prepare for questions that might be asked

Immediately Preceding Testing

_____ Make sure that facilities and seating arrangements are appropriate
_____ Organize any registration processes
_____ Organize testing materials

Administration of Instrument

_____ Perform any registration duties in the specified manner
_____ Distribute appropriate materials in the specified manner
_____ Administer instrument using the prescribed methods
_____ Adhere to specified time limits
_____ Note any behavioral observation of the client if individually administered
_____ Note any irregularities

After Completion of Testing

_____ Gather testing materials in the specified manner
_____ Follow specific directions for either scoring the instrument or sending the materials to be scored

Digital Download **Download at CengageBrain.com**

count the number of items that correspond to Holland's (1997) six personality types. The manual, however, warns that scoring errors do occur and that the counselor should take steps to minimize any errors. For example, some practitioners will double-check clients' scoring, whereas others will simply have clients recheck their addition.

Although hand scoring is becoming less prevalent, some clinical situations may involve hand scoring of an assessment. Even if the scoring is quite easy, there are times when it is therapeutically inappropriate to have the client score the instrument. In addition, with many instruments, hand scoring typically involves placing a template over the answer sheet and counting the answers that correspond to various scales. If hand scoring is used, counselors need to ensure that this often tedious process is completed with attention to detail and accuracy.

Computer scoring, when compared with hand scoring, is often more accurate, rapid, and thorough. Computers are not biased and will not unintentionally consider factors such as race or gender in scoring. In addition, computers can be programmed to complete

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

complex scoring procedures that would take humans weeks to perform. Furthermore, computers can be programmed not only to score but also to interpret the results. Computer scoring, however, is not always error free. Humans program the computers to perform the scoring, and errors in programming can result in problems. To ensure accurate scoring, the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) call for test-scoring services to document the procedures and accuracy of the algorithms they use. In addition, it calls for scoring services to monitor their scorings and correct any errors that might occur. This mandate, however, does not guarantee that all computerized scoring programs are sound, because individuals who lack sufficient training could write a program that improperly scores and interprets instruments. Before using a computerized scoring service, practitioners should investigate the integrity of that service and the steps that were taken to develop the scoring program. Another problem with computer scoring and interpreting relates to the ease with which assessment results can be generated. Because of this, some organizations might be tempted to skimp on staff training on an assessment device, or they might allow assessments to be conducted without appropriate professional oversight.

Not all assessments are scored by computers, and some instruments require clinician judgment as a part of the scoring process. For example, with some individual intelligence or general ability tests, the examiner needs to make some professional judgments in scoring and interpretation. The manuals for these instruments attempt to reduce the ambiguity in these judgments by providing explicit scoring instructions. In the personality area, many of the projective techniques involve professional judgments. For some projective techniques, there are systematic methods for scoring the client's responses. Another area in which scoring is more unstructured is with performance or authentic assessment.

LO3 An area of assessment in which scoring is a particularly salient issue concerns performance and authentic assessment. The terms **performance assessment** and **authentic assessment** are often used interchangeably, but there is a slight distinction between the two (Oosterhof, 2000). Both performance and authentic assessments are typically associated with testing that goes beyond paper-and-pencil tests to assessment that more closely approximates the skill being measured. All authentic assessments are performance assessments, but the reverse is not always true. Authentic assessments involve the performance of "real," or authentic, applications rather than proxies, or estimators, of actual learning. Authentic and performance assessments are typically associated with achievement testing. Performance assessment, however, has its roots in industrial and organizational psychology (Bray, 1982), where individuals were given activities such as the "in-basket technique." With these types of assessments, multiple-choice items are typically avoided; instead, performance on open-ended tasks is evaluated. Because many of the same strengths and limitations in scoring apply to both performance and authentic assessments, this portion of the book will simply use the term *performance assessments* with the intent of also encompassing authentic assessment.

One of the goals of these types of assessments is to see if knowledge can be applied. For example, instead of giving a multiple-choice test on the rules of grammar, a performance test would involve writing a grammatically correct business letter. Another example of performance assessment is observing a student complete a lab experiment. Performance assessments seek to assess complex learning and processes by assessing both the product and the process. Thus, they involve observation and judgments about how the individual performed the tasks (e.g., the process of performing a lab experiment) and the product produced (e.g., the results of the experiment). Performance assessments can sometimes assess the more abstract learning skills that traditional multiple-choice tests cannot adequately measure. There are, however, many difficulties associated with scoring these assessments.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The reason authentic and performance assessment is discussed here concerns issues of scoring. Objectivity in scoring becomes more difficult with the open-ended methods that are typically used in performance assessment. Sometimes the criteria for evaluating these measures are not adequately defined, and, thus, other factors can influence the judgments that are made. Particularly when process rather than product is being assessed, evaluators may fail to observe some important behaviors or will attend to behaviors that are unrelated to the skills being assessed. Because subjectivity is often part of performance assessment, examiners need to attend to multicultural issues in scoring. Evaluators' biases can sometimes seep into their observations and evaluations. Banks (2012) suggests that performance assessments can most appropriately be scored by using a rubric, evaluation checklists, or a common evaluation instrument. A rubric is different from a checklist in that a rubric contains a written description of each evaluative statement; whereas, with a checklist does not describe the needed behaviors in detail. A common evaluation instrument is designed by the evaluator and it typically provides descriptors for each point assigned. The common evaluation instrument is used for all students where points are typically circled and the accompanying information describes the student's performance.

According to Salend (2009) in evaluating educational skills, there are two types of instructional rubrics: holistic and analytic rubrics. Holistic rubrics involve rating the quality of an entire assignment and are frequently used for comprehensive performance assessments that are related to district, state, or national standards. Conversely, analytic rubrics are typically created by individual teachers to evaluate student assignments within a class. More state and national educational assessments are incorporating performance assessments, particularly in the area of writing. These holistic rubrics, therefore, need to be very detailed as they involve multiple raters of those essays. In cases of state and national performance assessment, the evaluators usually receive training and use very detailed scoring rubrics in order to minimize rater differences. For large-scale assessments, test publishers will evaluate and publish information regarding interrater reliability; one would want these reliability coefficients to be high when these tests are being used to make important decisions.

Communicating Results

Once an instrument is scored, counselors often have the responsibility for interpreting the results based on the validation evidence and communicating the assessment results to clients, parents, or other individuals. Communicating the results to clients is often one of the most important aspects of the assessment process. Although counselors are trained in communication skills, the interpretation of assessment results requires specialized knowledge and competencies. The communication of results is an important process because if this information is not communicated effectively, clients or others may misperceive the assessment results. For example, a common problem is confusing the percentile with the percentage correct. I once had a client who was initially upset with her performance at the 50th percentile. When she saw her score on the mathematical section of a standardized achievement test, she began to weep and said that she had never been good in math. As I explained that her performance was in the average range, she looked puzzled and questioned my interpretation. As the discussion progressed, I began to see that she believed the score indicated she had gotten 50% of the items correct even though I was using terms such as the *50th percentile.* Once the misunderstanding was corrected, the client was pleased with her performance and seemed to feel more efficacious about her mathematical skills.

The communication of testing or assessment results should not be perceived as a discrete activity but should be interwoven into the counseling. The reason for using any

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

informal or formal assessment is to answer questions that either the counselor or the client has. Thus, reporting results should be directly connected to the counseling's focus. Counselors need to be knowledgeable about the instrument and the client's specific results so that the focus during the interpretation of results can be on the client's questions and reactions rather than on what the scores mean.

Research Related to Communication of Results

Surprisingly, there is little research related to how practitioners can best communicate results to clients. Much of the research related to the benefits of clients receiving test information is in the area of career counseling. A number of studies conducted before 1985 related to test interpretation and communicating results from career assessments (Whiston, 2002). In a meta-analysis that included career assessment studies, Oliver and Spokane (1988) found an effect size of .62 for individual test interpretation and an effect size of .76 for group test interpretation. This means that those who received the test interpretation were more than two-thirds of a standard deviation higher on the outcome measures used than those who did not receive the interpretations.

Concerning assessment in personal counseling, Goodyear (1990) also noted the lack of research but concluded tentatively that those who receive test interpretations, regardless of the format, do experience greater gains in counseling than those who do not receive an interpretation. Related to working with couples, Worthington et al. (1995) found that married couples gained more from assessment feedback as compared with those who only completed the questionnaire.

Jones and Gelso (1988) found that individuals perceived interpretations that were tentative as being more helpful than interpretations that were more absolute. These individuals also expressed more willingness to see a counselor who used a tentative interpretation approach over one whose approach was more absolute.

Counselors sometimes want to know whether it is better to interpret results individually or in groups. Interestingly, the research is inconclusive in this area. However, when given a preference, clients *prefer* to receive results individually. Furthermore, clients report more satisfaction, clarity, and helpfulness when the interpretation is performed individually as compared with other formats, such as groups (Goodyear, 1990). Benziman and Toder (1993) examined whether children in mental health settings would prefer to receive assessment feedback from the examiner or therapist as compared with receiving the information from their parents. In this study, the children voiced a strong preference for receiving assessment results from their parents.

Although clients may prefer individual interpretation, some organizations may need to consider group interpretation because of the cost. Krivasty and Magoon (1976) found that individual interpretation was six times more expensive than group interpretation. Some counselors (e.g., school counselors) may want to consider combining modes of interpretation, such as providing group interpretation for general interpretive information and then following up with individual sessions to discuss specific results.

An area that has garnered some recent research concerns therapeutic assessment (Finn, 2007). Researchers (Finn & Tonsager, 1992; Newman & Greenway, 1997) found that providing MMPI-2 feedback increases clients' self-esteem and reduces levels of distress. In particular, they found clients reported better outcomes when counselors communicated MMPI-2 results using a collaborative model. Finn (2007) developed a model of therapeutic assessment. Rather than simply gathering information to better understand the client, the goal in therapeutic assessment is to make the assessment experience a positive one and to help create positive change for the client. Much of therapeutic assessment is geared for psychologists who regularly conduct batteries of psychological assessments, but some of it has applicability to assessment in counseling. Finn suggested that assessment feedback be given in a specific

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

order. First the counselor would start with Level 1 findings, which are results that map onto the way the clients already think about themselves. Next, the counselor would move to Level 2 findings, which reframe or amplify clients' ways of viewing themselves. Lastly, if the process is going well, the counselor moves to Level 3 findings in which the assessment results are more discrepant or conflict with the usual way the clients view themselves. In therapeutic assessment, Finn (2007) stressed the need for a collaborative environment.

In terms of therapeutic assessment's focus on a collaborative style of test interpretation, Ackerman, Hilsenroth, Baity, and Blagys (2000) found that clients whose initial sessions involved collaborative style were more likely to begin therapy than those who received the more traditional information gathering approach. Hilsenroth, Peters, and Ackerman (2004) found that those who received a collaborative approach to assessment in the first session rated the therapeutic alliance as better than those who used an information gathering approach. Furthermore, those who received the collaborative approach to feedback had as high or higher ratings of the therapeutic alliance throughout the entire psychotherapeutic process. Hilsenroth, Ackerman, Clemence, Strassle, and Handler (2002) also noted that the clients of therapists who received training regarding therapeutic assessment rated the therapeutic alliance higher in the initial sessions of psychotherapy than clients of therapists who did not receive training.

Poston and Hanson (2010) conducted a meta-analysis of 17 studies that examine assessment as a therapeutic intervention. As you will remember from Chapter 4, meta-analysis is a statistical technique in which researchers can combine the results of many studies. Poston and Hanson found an effect size of .42 when they compared the results of those who received a therapeutic assessment to those who received no assessment or a more traditional feedback session. This means that those who received the therapeutic assessment were almost a half of standard deviation above the comparison group on various outcomes. The Poston and Hanson study is an important contribution to the assessment literature and clearly indicates a more collaborative style when providing assessment feedback influences the therapeutic process, and contributes to clients achieving their therapeutic goals.

L04 The following are some general guidelines to consider when communicating results to clients.

Guidelines for Communicating Results to Clients

- Before any type of interpretation or communication of results to a client can be done, the counselor needs to be knowledgeable about the information contained in the manual. Validity information is the basis for the interpretation because this information informs practitioners on appropriate uses and interpretations. Furthermore, most manuals will explain the instrument's limitations as well as areas in which clinicians should proceed with caution.
- Interpreting assessment results is part of the counseling process, and, thus, typical counseling factors (e.g., counseling relationship, interaction between client and counselor) need to be considered. To communicate assessment results adequately, counselors need to use effective general counseling skills.
- Counselors need to develop multiple methods of explaining the results.
- Sometimes the communication of results involves terms such as *standard deviation*, *normal curve*, and *stanines*, which may not be familiar to many individuals. Preparing to explain the results in multiple ways is necessary because some clients will not understand their results if the counselor uses only standard psychometric terms. When appropriate, counselors should include visual aids because some people understand better by seeing the results rather than by hearing a report.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

- Whenever possible, the interpretation should involve descriptive terms rather than numerical scores. For example, if a client's score is at the 83rd percentile, a counselor could explain this as, "If a hundred people took this instrument, 83 of them would have a score at or below your score."
- It is better to provide a range of scores rather than just one score. The description of the results should be tied to the reason for the assessment. As an example, if a counselor gave a personality inventory to a couple, the counselor would need to explain the rationale for taking the personality inventory and how the results can be useful.
- The assessment results should not be presented as infallible predictions. When probability information is available, counselors should explain the results in terms of probabilities rather than certainties. There is, however, considerable research indicating that people attribute very different meanings to many of the probabilistic words—such as *unlikely, possible, probable,* and *very likely*—commonly used in assessment (Lichtenberg & Hummel, 1998). These findings indicate that counselors vary widely in their interpretations of probabilistic statements in interpretative reports. Lichtenberg and Hummel suggested that counselors should use numerical descriptors (e.g., there is a 95% chance) rather than verbal ones (e.g., there is a high probability) to convey more precise information.
- Although the results should contain detailed information, Jones and Gelso (1988) found that individuals perceived interpretations that were tentative as being more helpful than interpretations that were more absolute.
- The results should be discussed in context with other client information. Interpretation should never be done in isolation; rather, the interpretative process should explore how the assessment information conforms to other pertinent information. A competent interpretation involves not only analyzing the assessment results but also integrating diverse client information into a comprehensive understanding. Assessment results can be viewed as pieces of a puzzle, with other puzzle pieces being used to understand the total picture.
- When describing results, therapists should involve clients in the interpretation.
- Clients need to be involved in the process so that they can address questions about whether the results make sense or if the results fit with other information.
- Counselors should monitor client reaction to the information, which can best be accomplished by having the client provide feedback throughout the process.
- Asking the client for reactions early in the assessment process lays the foundation for an interactional interpretation.
- Any limitations of the assessment should be discussed in nontechnical terms. Some clients will be confused by comments such as, "The reliability coefficients are low." An explanation of the limitations needs to be geared toward the client's level of assessment sophistication. If there are questions related to the instrument's results in terms of the gender, racial background, or other characteristics of the client, then this also needs to be explained.
- The counselor should encourage the client to ask questions during the process.
- Occasionally, clients believe they will appear foolish to their counselors, so they may be reluctant to ask questions. To determine whether clients understand the assessment information, counselors can use incomplete sentences with the client filling in the information or completing the sentence. Counselors must ensure that clients do not leave confused or ill informed about the results.
- At the end of the conversation, it is often helpful to summarize the results, reiterating and stressing important points. This summarization process provides a second opportunity for the client to question inconsistent results.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L05 Guidelines for Communicating Results to Parents

The guidelines in the preceding list also apply to communicating assessment results to parents; however, some specific issues do arise when interpreting results with parents. Parents frequently receive achievement test information, but they may have little background in testing. In addition, schools sometimes send assessment results home even though the materials may not adequately explain the results. When interpreting the results, parents often seek help from a school counselor. Any counselor, school or otherwise, who works with a family should be prepared to interpret achievement and aptitude testing results. Table 7.1 contains a list of common questions that all counselors should be prepared to answer. Tharinger et al. (2008) have found that when parents and children are confident from the beginning they will receive a thorough explanation of the results that they are more likely to be open and cooperative. Therefore, when results are being interpreted, counselors should be prepared to answer parents' questions and able to explain the assessment results.

Increasingly, counselors are members of multidisciplinary teams that collaborate in the delivery of mental health services. As a part of a team, counselors may be responsible for counseling parents after a child has been diagnosed with a disorder. Examples of the more typical diagnoses a child may receive from a school psychologist or community practitioner are intellectual disability, communication disorders, autism spectrum disorder, attention-deficit/hyperactivity disorder (ADHD), specific learning disabilities, oppositional defiant disorder (ODD), and conduct disorder. In these cases, parents usually have received an explanation of the testing that was performed, but they are often overwhelmed with the amount of information presented. During their work with parents, counselors must have an understanding of the testing used and the symptoms of the disorder. If counselors are not part of the multidisciplinary team and do not have access to the testing results, they can gain access to the information by having parents sign an informed consent form to release the information to them.

Parents often struggle with the news that their child has a disorder, and a counselor can be instrumental in helping them adjust to the situation. Many parents in such a situation have experienced numerous difficulties and frustrations and need to interact with an individual who can empathize with their situation. A family's reaction to the diagnosis of a disability tends to go through five stages: impact, denial, grief, focusing outward, and closure (Fortier & Wanlass, 1984). To progress through the stages, a family often needs help

TABLE 7.1
Questions parents commonly ask counselors

_____ Why are the scores so low?
_____ What is an achievement test?
_____ Who was in the comparison group?
_____ What does criterion-referenced mean?
_____ What do these results mean for the future?
_____ What is a general abilities test?
_____ Who will see these results?
_____ What are percentile scores?
_____ Why do you test so much?
_____ What are stanine scores?
_____ Could my child have done better?
_____ Why aren't grade equivalent scores good?
_____ Where did the questions on the test come from?
_____ What can I do to help my child?

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

in understanding the child's condition. This may involve discussing the assessment report more than once. It is particularly important that the counselor be prepared to discuss the results by using a variety of techniques. The counselor may also help the parents focus on the child's abilities and not just on his or her disabilities—the entire family can benefit from understanding that the disability is only one aspect of a child's life. Sattler and Garro (2014) also encouraged clinicians to emphasize the active role that parents and their child can play in coping with the disability. The counselor should attempt to frame the family's response into an active, coping approach rather than a succumbing, victim framework.

Focusing on coping methods does not mean that the counselor should ignore the negative emotions a family is feeling. It is often very difficult for parents to hear that their child is, for example, autistic. Both the parents and the child often experience feelings of guilt and think that they somehow caused the problems identified in the testing. Frequently, there are also feelings of sorrow and loss. Even parents who have been aware of developmental delays and other problems for a period of time tend to experience feelings of loss when confronted with the actual results of testing. In general, these feelings are associated with the loss of the "perfect child" that most parents fantasize about at some point. Counselors, therefore, need to provide a therapeutic environment in which issues related to loss, anger, and frustration can be openly addressed.

In addressing negative reactions to the assessment results, the counselor needs to monitor the parents' responses so that the child does not internalize their negative reactions. Sometimes, a child will interpret the parents' disappointment in the testing result as disappointment in the child. The child may also be concerned about rejection by peers. Often when psychodiagnostic evaluation is completed, educational interventions are implemented (e.g., the child goes to a resource room or receives other special education services). The child may fear the reactions of other children as they become aware of the situation. Hence, this may be a time when the child is particularly vulnerable and, at the same time, is unsure of his or her parents' level of affection.

Psychological Reports

Many times, assessment results are communicated in written form, which is typically called a **psychological report** or psychoeducational report. This is particularly the case when psychologists, using a battery of tests, conduct a psychological evaluation. For example, a neuropsychologist may have conducted a battery of tests to assess an individual after a traumatic brain injury, or a school psychologist may have given a child a set of tests to determine if there is a learning disability. In most traditional positions, counselors will not be writing these extensive psychological reports but will use these reports in their work with clients. For instance, a counselor could be assigned to assist the individual with the traumatic brain injury cope with his injuries. To provide effective counseling, this counselor would need to be able to understand the neuropsychologist's psychological report. Likewise, school counselors often work closely with school psychologists and, thus, they must also be able to understand the psychological reports. Although there may be occasions when a counselor is called on to write a psychological report, the focus of the following section is more on counselors evaluating others' reports and applying the information in counseling interventions. A psychological report completed by a skillful clinician can provide an abundance of clinical information. However, psychological reports can vary in quality; therefore, counselors need to be able to evaluate the quality of the report before implementing interventions based on its content. (A detailed discussion on the writing of psychological reports is beyond the scope of this book; counselors who need these skills are directed to Groth-Marnat and Davis [2013] and Lichtenberger, Mather, Kaufman, and Kaufman [2004].) In fact, Groth-Marnat and Davis have software designed to help professionals write psychological reports.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 7.2
Typical sections in a psychological report

Identifying Information
Reason for Referral
Background Information
Behavioral Observation
Test Results and Interpretation
Diagnostic Impressions and Summary

The purpose of psychological reports is generally to disseminate assessment information to other professionals or sometimes clients or parents. For example, a report written by a school psychologist may focus on intelligence and achievement testing, with the interpretation directed toward teachers and how to improve their instruction. Many times, the psychological report is not geared exclusively toward the client, and the content may be difficult for some clients to understand because the information is somewhat technical in nature and often focuses on diagnostic and clinical decisions. Lichtenberger et al. (2004), however, warn that the client could read the report so the writing should be clear and readily understandable. A well-written report will not merely report test results but will also incorporate a wide array of relevant information. Thus, in evaluating psychological reports, counselors should expect a comprehensive overview of the client and an interpretation of the assessment results in a contextual manner. A psychological report should be carefully crafted with attention to detail. The caliber of the report can be initially evaluated by examining typographical errors, use of vague psychological jargon, careless mistakes, and lack of detail.

LO6 ***Common Areas in a Psychological Report.*** A counselor can expect a psychological report to contain certain information. Table 7.2 lists the typical sections included in a psychological report suggested by Lichtenberger et al. (2004). Because there are no established guidelines for writing a report, not all reports will follow this outline. A quality report, however, will include all of these major areas. Many reports will have a heading, such as "Psychological Evaluation," "Psychoeducational Assessment," or "Neuorpsychological Evaluation." Some such headings will also include the phrase "for confidential use only" to reiterate the confidential basis of the information contained in the report.

LO7 ***Identifying Information.*** At the beginning of a psychological report, there is usually some identifying information about the examinee. This section of the report provides some demographic information on the client. It is often done in outline format and includes the following information:

Client's name
Date of examination
Date of birth
Chronological age
Date of report
Grade or occupation (if applicable)
Examiner's name

Other possible information includes the individual's gender, the teacher's name, or the parents' names. All of this is important information and is necessary introductory information to the reader of the report.

Reason for Referral. After the identifying information, there should be heading entitled "Reason for Referral." This section of a psychological report typically addresses the reason

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

for the referral for testing. Some practitioners will use the label "Reason for Counseling" in order to broaden the focus. It should be noted that the term *psychological report* is typically associated with the reporting of assessment information. The information customarily included in this section is the name and position of the referral source (e.g., teacher); the reason for the referral; and a brief summary of the symptoms, behaviors, or circumstances that led to the referral. The reason for the referral should guide the evaluator in the selection of instruments or procedures. In evaluating a psychological report, the practitioner should find a direct connection between the instruments or procedures used and the reasons for the referral. Furthermore, there should be a direct connection between the referral questions and the summary at the end of the report.

Sometimes after the referral question section, you will find another section of the report with the title "Tests Used" or "Evaluation Procedure." This section is a simple listing of the assessments used. An example of this section would be Wechsler Intelligence Scale of Children-V, Connor's Teacher Rating Scale Revised-Long Form, and classroom observations.

Background Information. This section is critical for many readers because it provides a context in which to interpret the assessment results. The background information portion of the report should provide the reader with an overview of relevant information regarding the client. This section should also provide sufficient detail about the client's past so that the reader can understand the current issues and concerns. The information in this section may come from interviews with the client; interviews with others (e.g., a teacher or parents if the client is a minor); or past educational, counseling, or health records. The focus of the background review will vary depending on the reasons for the assessment, but it should include significant past events related to the assessment purpose. Educational history is often relevant, as well as the current family situation. Notable physical and mental health events in the client's life should be discussed. If other assessments have been performed, there should be a summary of those assessment results and the progress made since the prior evaluations. For adult clients, this section should include a description of the current employment situation and any other past or present work-related information germane to the assessment.

Behavioral Observations. Practitioners unfamiliar with standard protocol may not include this important section of a psychological report. However, a well-written psychological report will describe what is observed during the assessment process. This information is critical to interpreting the results, for it addresses issues such as motivation, anxiety, concentration, and self-efficacy. The behavioral observations are reports of the client's behaviors while taking the instruments or during the interview. The behavioral observations are critical to understanding the test results as they reflect the person's behavior while taking the assessments. The behavioral observation section will usually include a brief description of physical appearance, client comments and reactions to the process, responses to different assessment activities, unusual mannerisms or behaviors, variations in activities, and changes in voice tone and facial expressions. Groth-Marnat and Davis (2013) label this section "Behavioral Observation/Mental Status," which implies that the examiner should include information about the examinee's mental status (see Chapter 8 for more information on mental status). If inferences are made about the client's behavior, then those inferences should be accompanied by descriptions of behaviors supporting the conclusions. For example, a report might include statements such as, "The client appeared somewhat tense and chewed her fingernails. She also reported that she was feeling anxious about the testing." Occasionally, the evaluator might use a formal behavioral observation tool such as the Student Observation System (SOS) from the *Behavior Assessment System for Children, Second Edition* (BASC-2).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Assessment Results and Interpretations. This is the heart of the psychological report and will be the longest section. Sattler (2008) suggested that this section be entitled "Assessment Results and Clinical Implications," which directs attention to using the results to assist the examinee. Groth-Marnat and Davis (2013) called this section of the report "Impressions and Interpretations." Topics included in this section are the assessment findings, the meaning of the results, and the clinical impressions. In the discussion of the assessment findings, the precise scores on every scale and subscale may not need to be presented. In some reports, you will see a separate section where they simply provide the test results, and this section is focused on clinical impressions. The focus should be on the pertinent findings and *interpreting* rather than simply reporting the findings. The report should include sufficient detail so that a reader unfamiliar with the specifics of the instrument can understand the meaning of the results. Names of the instruments and descriptions of the scores and scales will help the reader understand the results. Sometimes tables of results are embedded in the test results and interpretation section of the report and sometimes tables of results are included at the end of the report. The interpretation of the results will often include an exploration of the client's strengths and limitations. Furthermore, there should be a discussion of both consistent and inconsistent scores and results. The reader will also expect to find an explanation or possible hypotheses about any notable results. The psychometric qualities of the instrument are not directly reported in the report, but the interpretation needs to incorporate any psychometric limitations. For example, limited validation evidence should be stressed. The report should be organized in an easy-to-follow manner. Many practitioners organize their reports in a domain-by-domain framework. A summary is not needed at the end of the results and interpretation section because one will be provided later in the report.

In evaluating a psychological report, a counselor should consider the quality of information and how the report is written. First, the writing needs to be specific and concrete rather than vague and ambiguous. A reader with little background in assessment should be able to understand the report. A vague statement such as "Pat has low self-esteem" is problematic because it is not clear what is meant by "low," nor is it clear what is meant by "self-esteem." Although the report should be clearly written, it should not include declarative statements about people. For example, a report should not say, "Michael is an extrovert." Michael may only be portraying himself to be extroverted because he thinks that is more socially acceptable. Instead, a report writer should report, "Michael's answers on the Myers-Briggs Type Indicator *indicated* that Michael's preference is toward the extroverted function, which reflects a preference for . . . ." The focus should be on what the scores reflect or indicate, not on some specific label. A clinician cannot say Mary's IQ is 102 because the next time she takes an intelligence test her IQ may be 104. It would be more appropriate to present the results as, "Mary's scores on the [insert test name] reflect that her IQ is 102." A better strategy would be to use standard error of measurement and say, "Mary's performance on the [insert test name] indicates that 68% of the time Mary's IQ could be expected to fall between 99 and 105."

Diagnostic Impressions and Summary. Some psychological reports will include diagnostic impressions, whereas others will not and the heading will simply be Summary. Sometimes diagnostic or clinical impressions are included in this section. No matter where it is located, when a diagnosis is presented in a report, there should be sufficient evidence to support that diagnosis. Consistent findings from different sources provide the sort of documentation expected for a diagnosis.

The summary is a succinct summarization of the entire report, with a focus on the results and interpretation. The summary may vary in length from a few paragraphs to longer depending on the number of instruments that were administered. As one of my teachers

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

once said, "Write the summary as if that will be the only thing anyone reads." The summary provides an opportunity to reiterate and emphasize important results. It provides a concise overview of the entire psychological report and emphasizes the major assessment results. If there are inconsistencies in results, then the evaluator should resolve these inconsistencies as well as possible.

Sometimes the summary and recommendations will be combined into one ending section and sometimes recommendations are given a heading by itself. A psychological report should include recommendations particularly as they relate to the reasons for the referral. The recommendations are not just based on the test results, but should be derived from the entire report including the background information and the behavioral observations. Some recommendations may involve future evaluations and with all recommendations it should be easy to see the connection between previous sections of the report and the recommendations. The recommendations should be realistic, with consideration of the client's resources and situation. There also needs to be sufficient detail so that the recommendations can be easily implemented. Many times, this requires the clinician to have information about community resources.

Signature. The report writer's name, degree, and professional title or capacity should be typed at the end of the report. The writer then signs her or his name above the typewritten name. If the psychological report is written by a student or intern, then there will also be a signature of the supervisor.

In conclusion, a thorough psychological report written by a well-trained clinician can provide a wealth of information that can be used in treatment planning. Most credible reports will identify both strengths and limitations of the client and will include pertinent recommendations. A quality report will not only provide the assessment results but also integrate other information to produce a thorough report. Counselors who write psychological reports should ensure that their writing is clear and concise and that it accurately describes and interprets the client's assessment results. More often, counselors will be consumers of psychological reports and will need to analyze the report to determine the quality of the information.

Summary

This chapter examined instrument selection, administration, scoring, and communication of results. Selecting and using an inappropriate instrument can have a negative impact on clients; therefore, selection of either a formal or informal assessment tool should be done carefully. An instrument should not be used with a client until the counselor has thoroughly evaluated it, including the instrument's purpose, the procedures used in constructing the instrument, the appropriateness of the norming group or criterion, the reliability, and the validation evidence. Instruments should also be evaluated for fairness, particularly in multicultural counseling situations. The practitioner's qualifications should also be considered in determining whether to use an instrument. It is unethical for counselors to use an instrument for which they have not received the proper training and the ultimate responsibility for ethical use of an instrument rests with the test user.

Part of using an instrument appropriately is ensuring that the instrument was properly administered. Counselors often shoulder this responsibility, even in settings such as schools where achievement tests are actually administered by classroom teachers. The key to ethical instrument administration is understanding and following the administrative instructions. Most manuals provide instructions detailing the manner in which an

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

instrument should be administered. Counselors need to follow these guidelines and prepare for potential difficulties. Foresight and planning can often circumvent potential problems in instrument administration.

Accurate scoring must be ensured to guard against erroneous findings. For many instruments, computer scoring is becoming more prevalent than hand scoring. Some instruments involve intricate scoring procedures and can only be computer scored. With certain types of instruments, scoring involves professional judgments by the person scoring the results. An assessment area that has been of increasing interest is performance and authentic assessment. Although subjectivity can be problematic in these types of assessment, there are procedures that can enhance the scoring process.

Research findings reflect that clients who receive test interpretations, regardless of the format, experience greater gains in counseling than those who do not receive an interpretation. Communicating the results to clients is often one of the most important aspects of the assessment process. However, there is surprisingly little research related to the most effective methods for communicating assessment results. In general, it is more effective to integrate the communication of assessment results into the counseling process. Clients prefer to have results interpreted to them individually, but group interpretations also appear to be effective and are significantly cheaper. There are research findings that indicate interpreting assessment results can be done in a therapeutic manner.

Assessment results must be interpreted in context, and part of this process involves examining whether the results are consistent with other information about the client. In addition, to facilitate understanding, the practitioner should be prepared to explain the results by using multiple methods. Clinicians may also need to explain results to parents to assist them in understanding the results and in making informed decisions about their children. Sometimes, the assessment information that counselors are interpreting to clients is from psychological reports written by other professionals (e.g., school psychologists). Counselors need to be able to distinguish quality reports from those that are carelessly or haphazardly compiled. A sloppily written report may be indicative of imprecision and carelessness in the entire assessment process. Occasionally, counselors may be required to write psychological reports; they should seek training on report writing and on appropriate documentation and interpretation of assessment results before attempting to write a psychological report.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 8

Initial Assessment in Counseling

Learning Objectives

- L01 Describe information typically gathered in an initial intake interview
- L02 Explain the strengths and limitations of structured and unstructured interviews
- L03 Identify other strategies that can be used in initial assessments (e.g., checklists)
- L04 Determine an assessment for suicidal ideation
- L05 Discuss methods for assessing for depression
- L06 Describe methods for assessing for potential violence
- L07 Understand how to assess for potential child abuse
- L08 Explain the importance of assessing for substance use
- L09 Express how to conduct a mental status exam

You have just been hired as a counselor for a community mental health center. Your first client is Cal, a 56-year-old African American male. The center does not have a structured intake procedure and leaves the process up to the clinician. Because you have never conducted an initial intake interview at this location, you are unsure of how to maximize your initial time with Cal. This community mental health center has a reputation for working with clients who abuse substances. Therefore, you think you should explore substance use with Cal, although you are not sure how to ask appropriate questions that will get honest appraisals of substance use. Furthermore, you do not want to miss any significant issues or important problems. Your supervisor has indicated there are some brief assessments in a file drawer that you can use if you think it might be helpful. Because the session is only scheduled for 50 minutes, you will need to use your time wisely.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

How a counselor begins the counseling process has been found to be critical to the eventual effectiveness of the counseling. Many clients do not return for a second session, which may be related to counselors' skills in identifying client issues and building positive expectations about the counseling process based on that assessment. Researchers have found that clients' expectations about counseling before they enter have very little influence on eventual outcome; however, the expectations that clients build during those initial sessions have a significant influence on outcome (Clarkin & Levy, 2004). Building positive expectations, however, needs to be based on accurate assessment of client concerns. Some of us have had the experience of going to a medical doctor and feeling like the physician is not really attending to our reporting of our symptoms. This analogy applies to counseling because clients want a clinician who can assist them with their specific issues. Also, building a therapeutic relationship starts early in the counseling process. Beutler, Malik, Talebi, Fleming, and Moleiro (2004) found that it is the quality of the therapeutic relationship in the early sessions of psychotherapy, rather than later in the process, that significantly predicts long-term positive results. Hence, from the beginning counselors should consider the initial session as being critical to the effectiveness of the overall process. In addition, counseling tends to be more efficacious when the counselor is able to quickly conceptualize the relevant issues.

Clients often come to counseling with scattered information regarding current issues, problems, and circumstances. Sometimes the presenting concerns are somewhat vague and ill defined. The concerns are often presented in statements such as, "I am not getting along with my husband," "The kid won't listen," "I keep making the same dumb mistakes," "I can't stand my job anymore," "I just can't get motivated to do anything," or "My son never does his homework, and he lies." Counselors need to become skilled in taking these vague concerns and identifying relevant issues and problems. Cormier, Nurius, and Osborn (2009) suggested that problems are rarely related to one factor, are usually multidimensional, and typically occur in a social context. Therefore, counselors need finely tuned initial assessment skills to identify the complex interaction of factors related to the client's problems.

L01 Initial Intake Interview

Counselors often use interviewing skills to begin the counseling process and obtain relevant clinical information. Initial, or intake, interviews usually gather information about the range and scope of the concerns, pertinent details about the current situation, and background information relevant to the current problems. Nelson (1983) found that this type of interview was one of the most commonly used assessment strategies. Despite research findings that the interview is the most popular assessment strategy used, it is not always easy to conduct one effectively. Research evidence indicates that guidelines and training are needed for counselors to obtain accurate, valid information from an interview (Duley, Cancelli, Kratochwill, Bergan, & Meredith, 1983).

In Hill's (2014) model of helping or counseling, the first stage of the process is *exploration*. In this stage, the counselor seeks to establish rapport, develop a therapeutic relationship with the clients, encourage clients to tell their stories, help clients explore their thoughts and feelings, facilitate the arousal of client emotion, and learn about clients from their own perspective. An effective initial interview is a balance between gathering information and building a therapeutic relationship. In Hill's model, the first stage of exploration lays the foundation for the second stage, *insight*, which leads to the final stage of the client taking *action*. The initial assessment and information gathering need not have

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

a detrimental effect on the relationship, because a skilled clinician can gather pertinent information while beginning to establish a working alliance. Therefore, this discussion first addresses information customarily gathered in an initial interview and then proceeds with an overview of therapeutic skills for conducting that interview.

Information Gathered in the Initial Interview

The information gathered in an initial counseling interview varies somewhat depending on the setting (e.g., school or agency) and the services delivered by the organization. The initial interview often provides information that influences the treatment phase of counseling and clinical decision making such as What are the client's major issues? Can counseling be effective with this person? and What is the best treatment approach or modality? To gather sufficient information to make sound clinical decisions, the initial interview needs to address demographic information, client background information, medical information, the client's presenting concern, and other relevant information concerning difficulties the client is experiencing.

Demographic Information

Typically in mental health settings, demographic information is gathered through the intake paperwork of an agency. For school counselors, much of the demographic information already exists in the student's educational file. It is important, however, to ensure that the demographic information is current and correct. Accurate telephone numbers and emergency numbers can be critical in crisis situations. In addition, pertinent demographic information needs to be gathered to make clinical decisions. For example, it could be important for the practitioner to have information on the client's marital status, race or ethnicity, employment status, and occupation. Also, for cases in which clinicians need to report suspected child abuse, information on all individuals living in a household and where the children attend school could prove to be important.

Client Background Information

To facilitate counselors' understanding of the client, they typically need to gather background information. Some clinicians call this gathering of background information a **psychosocial interview**. Gathering background information is not a detailed exploration of the client's past but rather an overview of the client's background, particularly as it relates to the presenting problem. Table 8.1 includes the common areas that are addressed in gathering background information. Counselors should explain to the client their reasons for gathering this background information. All of this information does not need to be collected in every intake interview. For example, in some mental health centers it may be more efficient to gather some of this information by using an easy-to-complete form. In some settings (e.g., schools) or for specific issues (e.g., career counseling), the counselor may not need to gather extensive information about the client's background. Nevertheless, to develop an effective treatment strategy, counselors will need to gather or have access to background information that is germane to the client's issues and the types of services provided by the counseling facility.

Health and Medical History

As part of gathering a client's background information, clinicians must have knowledge about the client's medical conditions so that they can holistically understand the client. This exploration of health issues should include an analysis of current health issues, including a listing of current medications. Because counselors want to ensure that there is no physiological base to the client's concerns, it is often wise to refer a client

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 8.1
Client background information

1. **Identifying Information**
   Client name, address, phone number, age, gender, marital status, occupation, workplace or school, work telephone numbers, name of another person to contact in case of an emergency
2. **Purpose of Counseling/Presenting Problem**
   Presenting concern and detailed information on the problem/concern
3. **Physical Appearance**
   Clothing, grooming, height, weight
4. **Present Level of Functioning in Work/School, Relationships, and Leisure**
   Analysis of present level of functioning in work, relationships, and leisure activities; degree that presenting concern affects activities
5. **Medical History and Current Status**
   Present health, current health complaints, health history, treatments received for current complaints, date of last physical examination, current medications, other drugs (either OTC or "street drugs"), sleep patterns, appetite level or changes, exercise patterns
6. **Family Information**
   Current marital status, number and ages of children living at home, other people living at home, any violence or physical abuse, family-of-origin makeup, significant influences of family-of-origin on present problems, family history of psychiatric problems, family history of substance abuse
7. **Social/Developmental History**
   Significant developmental events that may influence the present, any irregularities in development, current social situation, religious affiliation and values
8. **Educational/Occupational History**
   Occupational background, educational background, reasons for terminating previous jobs, military background, overall satisfaction with current job, stresses related to occupation
9. **Cultural Influences**
   Race and/or ethnicity, socioeconomic status, sexual orientation, acculturation, experiences of discrimination, perceived internal and external barriers
10. **Past Counseling Experiences or Psychiatric History**
    Type of treatment, concerns addressed, length of treatment, types of outcome, positive results, unhelpful results, medications prescribed for emotional/psychological problems

to a physician for a physical examination. Even if a client has had a recent physical, it is important to gather medical information and explore the possible side effects of any medications the client is taking. Certainly, with many clients, a thorough assessment needs to include an analysis of drug and alcohol use. Under the umbrella of health issues, counselors should also gather information on past psychiatric illnesses and treatments.

Client's Presenting Concern(s)

Many clients enter counseling because of specific issues that are causing them difficulties. It is important for the counselor to understand the issues or problems that may have motivated the client to pursue counseling. Outcome research results indicate there is a connection between attending to a client's presenting concerns and an effective outcome. This process involves more than noting the presenting concern. It also includes gathering information on when the issue started as well as surrounding events that may be relevant. The counselor should also explore the frequency of the problem, the degree to which the problem is affecting the client's daily functioning, and methods the client has used to attempt to deal with the problem in the past.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Other Relevant Information

In discussing an initial interview, it is difficult to address all of the necessary information that counselors should seek because counselors work in diverse settings and with a variety of clients. Hence, the content of an initial interview will vary and counselors may need to gather additional information that will augment what has already been covered in this chapter. In considering other information, clinicians need to focus on gathering *relevant* information. Counselors are human and have their own interests, which may influence their approach to an initial interview. For example, I am very interested in issues related to individuals' work and often I find it interesting to hear about individuals' employment history and their analysis of how they ended up in their current occupation. For some clients, focusing on work issues is relevant; however, for other clients it is not. Therefore, clinicians need to consider whether the information they are gathering is relevant in defining the client's problems.

Defining the Client's Problem

Once the clinician has gathered information about the presenting concern and about the client's background, the initial interview should focus on defining the client's problem or problems. Problem assessment plays a critical role in counseling because it is crucial that the counselor have a clear understanding of the presenting issues and concerns. Mohr (1995) found that one of the major reasons counseling can result in a negative outcome is that clinicians underestimate the severity of the client's problems. Many clients seek counseling for multiple issues rather than one specific problem. Therefore, it is important that counselors investigate the range of issues. Once the counselor has identified the dominant concerns or problems, the counselor should then explore each issue in detail. The following are topics to address in the examination of significant problems.

1. *Explore each significant problem from multiple perspectives.* Counselors should attempt to view the problem or concern from diverse vantage points. They can accomplish this by having the client describe the problem in terms of affective, behavioral, cognitive, and relational perspectives. If a counselor is working with a family, then having each member describe the problem contributes to a multidimensional perspective. If a counselor is working with a single client, then asking the client how others may describe the problem can promote a better understanding of the problem.
2. *Gather specific information on each major problem.* The counselor should encourage the client to describe the problem(s) in detail. Included in this description should be information on when the problem began, the history of the problem, what events were occurring when the problem arose, and what factors or people contributed to the situation. Information on the antecedents and consequences of the problem can be particularly helpful later when selecting treatment interventions.
3. *Assess each problem's intensity.* Counselors need to assess the degree to which the problem or issues are affecting the client. The problem's severity and the frequency or duration of problematic behaviors should be explored. One example of doing this is to have clients rate the severity of the problem on a scale from 1 to 10, which can provide an indication of the degree of negative impact. In certain situations, such as panic attacks, the counselor may have the client chart or record when and where the problematic behaviors occur.
4. *Assess the degree to which the client believes each problem is changeable.* When assessing client problems, it is important to explore the degree to which the client believes the problems are malleable or immutable. Assessment in this area will be helpful later in determining interventions and possible homework assignments.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

5. *Identify methods the client has previously used to solve the problem.* Counselors need to explore methods the client has already used to address the problem. Doing so provides insight into the client and avoids duplicating efforts that have already been attempted.

Once the counselor has assessed each significant problem the client is experiencing, the next step is to prioritize them. If the counselor addresses all of the problems at once, the counseling process will likely be stymied. Prioritizing the problems is considered part of the assessment process and can serve as a transition into the treatment phase. Determining which problems to address first in counseling depends on the situation, the client, and the counseling relationship.

Interviewing Skills and Techniques

A sound interview is accomplished by gathering information in a therapeutic manner. Therefore, counselors need to attend to both the content of the initial interview and to the processes they use in gathering that information. Because typical content has been covered already in this chapter, I want now to focus on the process of interviewing and the skills and techniques involved in the initial interview. The initial interview is the first contact with the client, and first impressions do have an influence on counseling outcome. Counselors should consider the degree to which they appear to be credible to the client in terms of being seen as trustworthy, expert, and attractive. LaCrosse (1980) found that initial client perceptions of counselors' expertness accounted for 31% of the variance in favorable therapeutic outcomes. Clients who perceive their counselors as being trustworthy, expert, and attractive helpers have better outcomes in counseling than clients who do not have these perceptions of their counselors (Heppner & Claiborn, 1989).

Numerous people have written about the importance of using effective communication skills in initial interviews (Cormier et al., 2009; Chang, Scott, & Decker, 2012). Communication skills are needed to communicate to the clients that they are being heard and understood in their responses to the interview questions. Open-ended questions are often used in interviewing because they are less likely to lead or unduly influence the client. In closed-ended questions, there is a set of "closed" responses from which the client can select an answer, for example answering either "yes" or "no" to questions such as "Did you come to counseling because you are feeling depressed?" Open-ended questions require a more elaborate answer that the client has to construct. Examples of open-ended questions are "What brought you to counseling?" "Can you describe for me the problems that you are having now?" "What situations are not going as well as you would like them to go?" With interviews that do not have a structured format, counselors may use other communication skills to draw out the client. Commonly used techniques are paraphrasing, clarifying, reflecting, interpreting, and summarizing (Okun & Kantrowitz, 2007). Furthermore, the counselor's vocabulary should be adjusted to correspond to the client's educational and social background.

Both a counselor's verbal and nonverbal behaviors impact the effectiveness of the initial interview. The climate the counselor builds during the interviewing process influences the degree to which clients are willing to disclose personal information. Asking questions is not the counselor's only verbal behavior during the initial session; restatement and reflection of feeling can both solicit additional information and facilitate a therapeutic environment. Table 8.2 includes some general guidelines for conducting initial counseling interviews. It should be remembered that clients respond differently to various therapeutic styles and there is not one correct manner of conducting an initial interview.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 8.2
Common guidelines for conducting an initial interview

- Assure the client of confidentiality (clearly state any exceptions, such as child abuse or harm to self or others).
- Word questions in an open-ended format rather than a closed format.
- Avoid leading questions.
- Ask questions in an accepting and respectful manner.
- Consider the client's cultural background.
- Listen attentively.
- Adjust your approach to the individual client (some clients are more comfortable with a more formal process; others prefer a more warm and friendly approach).
- Avoid "chatting" about less important topics.
- Encourage clients to openly express feelings, thoughts, and behaviors.
- If clients drift from pertinent topics, gently direct them back to the appropriate topics.
- Use voice tones that are warm and inviting, yet professional.
- Allow sufficient time, and do not rush clients to finish complex questions.
- Avoid psychological jargon.
- Vary your posture to avoid appearing static.

Interviewing Children

The previous suggestions on interviewing, such as establishing rapport, gearing the vocabulary toward the client's educational level, and asking questions in a warm professional manner, apply to children as well. Children, however, are sometimes more difficult to interview, so the counselor needs to adjust the initial interview to the child's developmental level. Starting with statements such as "I am going to talk with you for a little while to see if the two of us can figure out some ways to make your life better" or "Your mom told me some problems were going on, but I would like to hear what you think is going on" may help children understand the purpose of the initial interview.

It may also be helpful for counselors to provide an explanation of why they are asking questions at this stage in the counseling. For example, "I really want to understand, so I am going to ask you some questions to help me understand." Children, in particular, need to understand the purpose for the interview and counseling, because they have had little exposure even to media representations of counseling. Many children may see a counselor as an adult in a punitive role and will have difficulty trusting the counselor. Therefore, when attempting to identify problems with a child, it is important for counselors to define the limits of confidentiality.

Merrell (2007) contended that to interview children effectively, the interviewer must have knowledge of some basic developmental issues and be able to structure the interview so that it is developmentally appropriate. For children in early childhood, Merrell recommended (1) using a combination of open and closed questions; (2) not attempting to maintain total control of the interview; (3) reducing the complexity of the questions; (4) using toys, props, and manipulatives; and (5) establishing familiarity and rapport. Even though children are able to think more complexly as they move into middle childhood, Merrell suggested that the counselor (1) avoid abstract questions; (2) rely on familiar activities and settings; (3) not maintain constant eye contact; (4) provide contextual clues (e.g., pictures, examples); and (5) use physical props in context. In adolescence, children typically have the ability to reason abstractly and problem solve systematically so that the interview can be less concrete. With adolescents, Merrell recommended that the interviewer (1) consider the degree of emotional liability and stress, (2) avoid making judgments based solely on adult norms, and (3) show respect.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

If, during a counseling session, a child becomes involved in disruptive play, begins crying, or misbehaves in some way, Greenspan and Greenspan (1981) recommended not stopping the behavior too quickly. They suggested that observing this behavior rather than stopping it may be insightful and that disciplining too quickly may negatively affect the interviewing process. On the other hand, counselors do not have to sit and watch children destroy their offices. Simple statements such as "Can I ask you not to do that?" "We don't allow behavior like that in the center," or "Let's talk for just 10 minutes, and then we will do something else" can be helpful. Periods of silence or times when the conversation halts are more common in interviewing children than with adults. Silence may be an indication that the child is thinking about what to say or is reflecting on a situation or problem. If the silence becomes extreme, then the counselor might intervene with a question such as "These questions seem difficult for you to answer. Can you tell me about that?" or "You seem a little quiet. What are you thinking about right now?" In assessing children and their problems, counselors need to be more patient and skillful during the assessment than they are with adults.

LO2 Structured and Unstructured Initial Interviews

One of the decisions a counselor needs to make about the initial session is whether to use a structured or an unstructured interview. Sometimes this decision will be made for the counselor, because some agencies require that a structured interview be performed with each new client. In a **structured interview**, counselors ask an established set of questions in the same manner and sequence for each client. In an **unstructured interview**, counselors may have an idea about possible items to address but gear the interview in a unique manner, depending on the client's needs. A **semistructured interview** is a combination of the structured and the unstructured interview formats; certain questions are always asked, but there is room for exploration and additional queries. There are advantages and disadvantages to all three methods of interviewing.

The major advantage of the structured interview is that it is more reliable (Aiken & Groth-Marnat, 2005). Reliability concerns the proportion of error, which counselors want to minimize. Furthermore, as Chapter 4 addressed, reliability is a prerequisite to validity. Thus, a structured interview may possibly be more valid than an unstructured interview, although this is not guaranteed. Some practitioners believe that structured interviews have a negative impact on the counseling relationship because the client may feel interrogated. This is not necessarily true, because the interviewer's style has a significant influence on the counseling relationship and the degree of comfort the client feels in disclosing information. The soundness of a structured interview is intimately linked to the questions included in the interview and to the counselor's ability to elicit crucial information. It is difficult to assemble a structured interview that responds to the array of problems and situations faced by clients. Therefore, with structured interviews, it is important for counselors to include a question at the end to encourage the client to disclose any important information that may not have been addressed.

The advantage of unstructured interviews is that they can be easily adapted to respond to the unique needs of the client. Unstructured interviews, however, are less reliable and more prone to error than structured interviews. In my supervision of counseling students, I have seen unstructured interviews in which the initial interviews focused entirely on a relatively minor issue and the counselors neglected to gather adequate information. In unstructured interviews, gathering information tends to take longer as compared with well-developed structured interviews. The decision to use either a structured or an unstructured interview should be based on the interview's purpose. For example, if the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

purpose of the initial interview is to screen clients to determine if they are appropriate for a clinic, then a structured interview that reduces the amount of error is probably better. If, on the other hand, the purpose is to better understand the specifics of an individual client, then an unstructured interview may allow for more individualized exploration. The unstructured interview requires much more skill on the part of the interviewer than does the structured interview.

The following dialogue is an example of an unstructured initial interview with Anna, a 16-year-old female client who has entered counseling at her mother's insistence.

Counselor: Hello, Anna. I understand from your paperwork that your mom wants you to be in counseling, but I would like to know from you what you think is going on.

Anna: Oh, I don't know. My mom is really controlling and wants me to be perfect. She is the one who should be in counseling.

Counselor: What kinds of things is your mom controlling about?

Anna: She is controlling about my school work and grades mainly. She wants me to be "clone-child" and get all As.

Counselor: What are your grades? Tell me about your last report card.

Anna: Well, they weren't as good as they used to be; but still, I only had two Ds and one F.

Counselor: What were your grades in the past?

Anna: Oh, I used to get really good grades, like mostly As and Bs, and my mom keeps harping that I should still be getting those kind of grades again.

Counselor: When was the last time you got grades that didn't upset your mom?

Anna: Oh, I don't know. I was still on the honor roll when I was first in high school, but I haven't been for, maybe, a year.

Counselor: Has anything changed for you?

Anna: What do you mean?

Counselor: Is it a different school? Are you taking different classes? Are your friends different? Do you feel differently about school?

Anna: I do feel differently about school—I just hate to go.

Counselor: Tell me more about this hating to go.

Anna: I just hate going to school. The kids there are so into themselves and who's in their group that they never even talk to anyone else. The teachers just want to make your life hard and point out what you did wrong. I just never want to go in the mornings.

Counselor: So you hate to go . . .

Anna: I hate it when I'm there, and I'm not learning anything worthwhile. I learn more when I stay home and read or watch TV. I don't really need to go to school, and it's better when I just stay home.

Counselor: Do you stay home from school sometimes?

Anna: Well, sometimes.

Counselor: How often, on average, do you stay home?

Anna: A couple of times a week.

Counselor: In the past two weeks, how many times have you missed school?

Anna: Lots.

Counselor: I'm not sure what that means. I'm not going to give you a lecture, so tell me how many days you went to school in the past two weeks.

Anna: Twice.

Counselor: When you don't go to school, what do you do? Tell me about a typical day.

Anna: Are you going to tell my parents?

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Counselor: No, unless you're doing something to hurt yourself or others. On the other hand, your parents could be charged with something called educational neglect, so that is something we need to think about.

Anna: My parents already sorta know because the school has contacted them and everything. I get up in the morning and act like I'm going to school, but my parents usually leave for work before I need to leave for school. I just wait until they leave and stay at home. My mom has kinda gotten wise to this lately and has been going in to work late. On those days, I go sit in the shed behind our house and wait until she leaves. It is just so hard for me to go to school.

Counselor: Tell me about how you feel when you're at home and how you feel when you're at school—I want to understand this better.

Anna: When I am at school, I just feel bad. No one talks to me; I don't have anyone to eat lunch with. The kids just look through me, or there's this group in my first period class that makes these stupid remarks when I do come to class. At home, I don't have to put up with that, I can just be myself and watch TV. I mean, sometimes, I'd like to be at school and be popular and everything, but it's so hard. I don't have any friends. The friends I used to have don't want me in their group—they made that very clear.

Counselor: I'm getting this feeling that being at school is quite painful for you. On a scale of 1 to 10, with 1 being you are feeling very happy and comfortable and 10 being the most awful pain, what number on the scale would represent your feelings while at school?

Anna: About an 8 most of the time that I'm there. It's most difficult when I first get there in the mornings and at lunchtime.

Counselor: Not that I'm saying your parents understand it very well, but if I asked them about this problem of you not going to school, what would they say?

Anna: They would say I'm lazy and that I'm being childish. They tell me to just make some new friends.

Counselor: I imagine you've tried to make new friends; tell me about some of the ways you've tried.

Anna: I try to go up and talk to people, but I'm not very good at it. I'm just really shy, and it's hard for me. When I start to talk to people, I sound stupid, so I stop.

Counselor: I know it's difficult to make new friends, and some people find it easier than others. On the other hand, I think we can work together and find some ways that you can be comfortable with to make some friends. It sounds like the lack of friends is what keeps you from going to school—am I hearing that correctly?

Anna: Yes, you're hearing it right. If I had some friends, especially to eat lunch with, then I could go to school.

Counselor: How about we meet together for three sessions for the next three weeks and see where we are and if we're making progress on making friends. Could you do that?

Anna: I could do that.

Strengths and Limitations of Interviewing

As mentioned earlier, in terms of reliability, structured interviews are more reliable than unstructured ones. However, the quality of the questions asked influences the validity of an interview and what information can be inferred from the client's responses. Counselors should consider the validation evidence and the inferences they make about a client

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

based on the initial interview. The information gathered in an interview must be carefully analyzed. In addition, counselors should consider other evidence that may substantiate or challenge the clinical judgments being made. The strength of any type of interview is its ability to question the client about issues, problems, and personal information. The limitations of interviews usually relate to the lack of validation evidence and the influences of counselor subjectivity. As Dawes (1994) demonstrated, clinicians tend to be biased in decision making. Therefore, clinicians need to ensure that they gather and analyze the information gained from interviews by using objective and sound methods.

All clients will not respond in the same manner to a clinical interview. Certainly, gender and culture should be considered in determining whether to use this assessment strategy. Some cultures do not encourage immediate self-disclosure to someone outside of the family. Clients may have difficulty relating to questions and formulating a response. A counselor could perceive this difficulty as resistance when it may be more a matter of unfamiliarity and discomfort. Interviews are a good method for gathering initial information; however, they are not the only method. Counselors may consider supplementing their intake interviews with other information, such as checklists or rating scales, to provide information possibly not addressed in the interview. Furthermore, the information from checklists or rating scales may complement the interview and aid the counselor in identifying the more pressing and significant issues.

LO3 Other Strategies That Can Be Used in Initial Assessment

The counseling process frequently begins with the counselor gathering information about clients and their concerns. Most practitioners start this process by talking with clients and using a structured, semistructured, or unstructured interview. Some counselors, however, supplement the interview with other assessment tools. These tools may include informal assessment instruments (e.g., checklists or rating scales) that are incorporated into the initial paperwork. Other assessment devices may be more formal instruments that provide information on symptoms and issues.

Checklists

A checklist can be a relatively simple and cost-effective method for gathering initial information from clients. Usually in a checklist, individuals are instructed to mark the words or phrases in the list that apply to them. Checklists can be filled out by the client or by an observer (typically a parent or teacher). When multiple individuals complete the same checklist on a client, a counselor can acquire a rich body of therapeutic information from diverse perspectives. For example, clinicians working with children can expand on in-session observations by having children's parents and teachers complete a brief checklist.

Standardized Checklists. Counselors can construct their own checklists or use a number of existing standardized checklists. Some of the existing standardized checklists focus on clients reporting their current symptoms. *The Symptom Checklist-90—Revised* (SCL-90-R; Derogatis, 1994) and the *Brief Symptoms Inventory* (BSI; Derogatis, 1993) are two popular instruments that often are used at intake in many mental health settings. These instruments are related and are very similar; the main difference is that the SCL-90-R contains 90 symptoms, whereas the BSI contains 53. In these checklists, the client simply responds to the listed symptoms using a 5-point scale of distress, ranging from 0 (not at all) to 4 (extremely). The SCL-90-R takes about 15 minutes to complete; the BSI is less time consuming.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

There are two methods for using these two symptom inventories as valuable initial screening tools. First, a clinician can simply scan the results of these instruments and examine which symptoms are distressing to the client. Included in these measures are items such as thoughts of ending their life, feelings of worthlessness, feeling afraid in open spaces, and feeling that most people cannot be trusted. Alternatively, the SCL-90-R and the BSI can be scored by providing measures on the following nine scales: somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. There are also three composite scores that can be particularly helpful to counselors in determining whether to refer a client for psychiatric evaluation: the Global Severity Index, the Positive Symptom Total, and the Positive Symptom Distress Index. Client responses can be compared with nonpatient adults, nonpatient adolescents, psychiatric outpatients, and psychiatric inpatients. The internal consistency measures and test-retest coefficients for the SCL-90-R and the BSI are quite high over short time periods. Derogatis (1993) recommended that the results of either the SCL-90-R or the BSI are good for no more than seven days. Derogatis (2000) also developed an 18-item symptom inventory called the Brief Symptom Inventory-18 (BSI-18) for settings that want a quick assessment or short instrument, and the total score is designed to provide a Global Severity Index. In addition, the BSI-18 contains six items that assess the three subscales of Somatization, Depression, and Anxiety. There are many other commercially available checklists that can be used with adults.

A number of checklists have been developed for use with children and adolescents. A widely used set of checklists for children is the *Achenbach System of Empirically Based Assessment*, which has instruments for preschool children, school-age children, and young adults, which will be discussed in Chapter 13.

Informal Checklist. Some counseling centers and agencies have developed informal checklists to be used in their facilities. For an example of one of these brief checklists, see Table 8.3. Some of these agencies have performed studies on the reliability and validity of these instruments, but many have not. If there is no psychometric information available for a checklist, then the practitioner should proceed with caution. Although it is possible

TABLE 8.3
Example of an informal checklist

Personal Checklist

Please check the concerns that apply to you.

___Depression	___Emotional/physical/sexual abuse	___Career issues
___Anxiety	___Binge eating	___Academic concerns
___Stress	___Self-induced vomiting	___Financial pressures
___Anger control	___Relationship problems	___Multicultural issues
___Panic attacks	___Family conflict	___Sleep disturbance
___Fears/phobias	___Conflict with friends	___Physical complaints
___Grief/loss	___Conflict with people at work	___Obsessive thinking
___Suicidal thoughts	___Lack of relationship	___Alcohol/drug concerns
___Unwanted sexual experience	___Indecisiveness	___Lack of assertiveness
___Sexuality issue	___Low self-esteem	___Trauma/assault/accident

___Other (please specify): _____

Digital Download **Download at CengageBrain.com**

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

to gather some preliminary information from these informal checklists, counselors need to be cautious in using these results without reliability and validity information. The results of informal checklists that do not have psychometric information can be viewed only as possible problems; a counselor's clinical decisions should never rest on this information alone. In using informal checklists, clinicians should also be aware that some clients have a tendency to check numerous items whereas others check very few. Therefore, a client who has checked numerous problems should not necessarily be viewed as being more disturbed than a client who checked only a few.

Rating Scales

A rating scale is slightly more sophisticated than a checklist because it asks the client or an observer to provide some indication of the amount or degree of the problem, attitude, or personality trait being measured. Sometimes rating scales contain a numerical scale, with numbers assigned to different categories. For example, clients may be asked to rate their level of anxiety from 1 (not at all anxious) to 5 (extremely anxious). Another common rating scale is the semantic-differential, in which individuals choose between two bipolar terms (see Table 8.4 for an example).

When counselors interpret ratings, they should keep in mind a number of findings related to these measures. In general, people tend to respond in the middle of the ratings, which is called *central tendency error*. Individuals are less likely to rate themselves, or others, at either the low or the high extremes. Another finding pertains to *leniency error*, which concerns an individual's reluctance to assign unfavorable ratings. People are more likely to endorse the positive as compared with the negative. Finally, in terms of an individual's ratings of someone else, there is often a *halo effect*. Research indicates that people are

TABLE 8.4
Example of an informal rating scale

Personal Rating Scale

Please place an X on the point that most accurately reflects how you typically feel.

Calm	___ ___ ___ ___ ___	Nervous	Courageous	___ ___ ___ ___ ___	Afraid
Superior	___ ___ ___ ___ ___	Inferior	Optimistic	___ ___ ___ ___ ___	Pessimistic
Sad	___ ___ ___ ___ ___	Happy	Independent	___ ___ ___ ___ ___	Dependent
Indecisive	___ ___ ___ ___ ___	Decisive	Abusive	___ ___ ___ ___ ___	Victimized
Outgoing	___ ___ ___ ___ ___	Shy	Adventurous	___ ___ ___ ___ ___	Timid
Strong	___ ___ ___ ___ ___	Weak	Confident	___ ___ ___ ___ ___	Anxious
Fearful	___ ___ ___ ___ ___	Confident	Irrational	___ ___ ___ ___ ___	Rational
Compulsive	___ ___ ___ ___ ___	Lackadaisical	Energetic	___ ___ ___ ___ ___	Lethargic
Patient	___ ___ ___ ___ ___	Impatient	Confused	___ ___ ___ ___ ___	Insightful
Trusting	___ ___ ___ ___ ___	Suspicious	Honest	___ ___ ___ ___ ___	Shrewd
Aggressive	___ ___ ___ ___ ___	Submissive	Moral	___ ___ ___ ___ ___	Wicked
Relaxed	___ ___ ___ ___ ___	Tense	Impulsive	___ ___ ___ ___ ___	Cautious
Hostile	___ ___ ___ ___ ___	Gentle	Competitive	___ ___ ___ ___ ___	Cooperative
Sluggish	___ ___ ___ ___ ___	Active	Hopeless	___ ___ ___ ___ ___	Elated
Sick	___ ___ ___ ___ ___	Well	Even-paced	___ ___ ___ ___ ___	Erratic

Digital Download Download at CengageBrain.com

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

influenced by first impressions, and these impressions influence subsequent ratings. Thus, there can be either a positive halo or a negative halo effect.

Informal rating scales can be incorporated into the initial assessment of clients. For example, clients could be asked to rate on a 5- or a 7-point scale the severity of the presenting problem, the degree to which the problem is affecting their daily functioning, their commitment to addressing the problem, and the likelihood that the problem will change. These informal rating scales can be asked verbally or incorporated into an intake form.

There are also standardized rating scales, such as the *Conners* 3rd Edition (Conners 3; Conners, 2008). The Conners 3 are instruments designed for screening children ages 6 through 18, with the instrument being completed by the parent, teacher, or adolescent. The three versions of the scale (for parent, teacher, and adolescent) include both long and short forms. This rating scale includes a number of scales (e.g., general psychopathology, inattention). Furthermore, this instrument is often used in the diagnosis of attention-deficit/hyperactivity disorder.

In conclusion, a number of both formal and informal instruments can supplement the initial client information gathered in an intake interview. Some counselors incorporate these techniques routinely, whereas others periodically use checklists, rating scales, or other types of inventories, depending on the situation. Counselors can improve their effectiveness by using multiple methods for gathering information. No one source of information, such as a test, should be the sole basis of any decision. Multiple sources of information will provide a more comprehensive overview of clients and their problems. Sometimes, in the process of gathering information, a counselor will determine there is a need to assess the client more thoroughly on a specific problem. For example, during the initial intake interview, the counselor may see signs of depression and want to gather information specific to depression. Thus far, we have discussed general initial assessment, but counselors also need to consider methods for assessing specific client problems.

Assessing Specific Problems in the Initial Sessions

Certain problems should be given special consideration during counseling's initial stages because if these specific issues are overlooked, there could be negative results, even the death of a client. A significant example is suicide, for if a counselor does not assess a client's suicide potential, a client might attempt suicide when treatment might have prevented this attempt. Depression is significantly related to suicide and is a second factor that counselors should consider during the initial stage of counseling. Assessment of potential violence has become increasingly more important as there are shootings at schools and other public locations. Child abuse is another topic that may come up in the initial interview and it is critical that counselors handle these disclosures appropriately. A final topic that counselors should assess in the initial intake is substance abuse because of the psychological and physiological effects of psychoactive substances. Although clinicians will want to consider many other specific problem areas in the initial assessment stage, these are discussed in detail because of the potentially serious ramifications of misdiagnoses.

LO4 Assessment of Suicide Potential

In the initial session, a counselor needs to evaluate each client's potential for suicide. Neglecting to identify a client with suicidal ideation could result in an avoidable death. The probability of working with a potentially suicidal client is high. According to the Centers for Disease Control and Prevention (CDC, 2015) suicide is the tenth leading cause of death in the United States and accounted for 41,149 deaths in 2013. This means that, on average, 113 people commit suicide each day. Although many people are successful when they attempt suicide, many more people attempt suicide but are not successful. This is sometimes

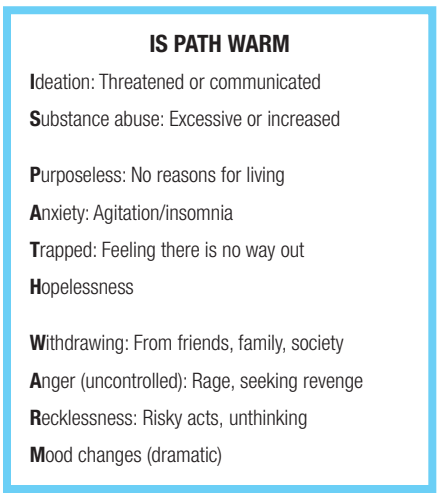
Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

called parasuicides and it is estimated that approximately 25 parasuicides occur for every completed suicide (Juhnke, Juhnke, & Hsieh, 2012). Rogers, Guelulette, Abbey-Hines, Carney, and Werth (2001) found that 71% of counselors had worked with individuals who had attempted suicide, and 28% of these practitioners had a client who had died by suicide. Predicting which clients might attempt suicide is very difficult because individuals' moods often vary greatly from day to day and even from hour to hour. A counselor may work with a client who appears to be at minimal risk; however, a few hours later something may happen to significantly alter the risk of suicide (e.g., the client's spouse asks for a divorce).

Granello (2010) contended that assessing clients for suicidal risk is one of the more challenging experiences a counselor faces. Granello also reminds the counselor that assessing risk of suicide is unique for each client although there are some common signs. In assessing for suicidality, the mnemonic device in Figure 8.1 can be helpful in remembering the signs of suicide. The I in IS PATH WARM stands for ideation and this concerns whether the client is having suicidal thoughts or ideations. As Stelmachers (1995) indicated, a client who verbalizes the thought of suicide is providing a clearly identifiable signal for the counselor to investigate the situation's lethality. Any suicidal ideation or thoughts should be followed up by a more intensive assessment even if the client indicates the thoughts are relatively minor. Many of the people who are contemplating suicide want to end their pain and may see suicide as the viable solution to ending that pain. Sometimes clients will use euphemisms or veiled threats, but it is important that counselors do not respond in a like manner and talk clearly with the client about his or her possible ideation (Granello, 2010). The next letter in the IS PATH WARM mnemonic is S, which represents substance abuse. Although there are many different definitions for substance abuse; in general, the counselor should be concerned if substance use is causing problem in the client's life. Of course, substance abuse by itself is not as strong of predictor of suicidality as some of the other predictors, but it is an issue that can significantly contribute to suicide attempts and, therefore, it is something that counselors should monitor. Increased use of alcohol or drugs can also be problematic as substance use can cloud decision making and contribute to suicidal ideation.

The P in IS PATH WARM stands for purposeless and the counselor should be particularly concerned if a client indicates that he or she has no reason for continuing or living. Although the A in the mnemonic device stands for anxiety, clinicians should be concerned if the client has psychological vulnerabilities or mental disorders. As Figure 8.1 states that

FIGURE 8.1
Mnemonic for red-flags for suicidal behavior


IS PATH WARM
Ideation: Threatened or communicated
Substance abuse: Excessive or increased
Purposeless: No reasons for living
Anxiety: Agitation/insomnia
Trapped: Feeling there is no way out
Hopelessness
Withdrawing: From friends, family, society
Anger (uncontrolled): Rage, seeking revenge
Recklessness: Risky acts, unthinking
Mood changes (dramatic)

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

A is related to anxiety, the counselor should also listen for symptoms of agitation and insomnia. The T stands for trapped and particularly feeling like there are no alternatives or there is no way out. Sometimes from certain client's perspective, suicide can be seen as the only alternative as they feel trapped within their situation and feel like suicide is the only solution for getting out. The H stands for hopelessness, which should be considered as a significant predictor of possible suicidal intent. The degree of hopelessness should be explored with clients who verbalize some hopeless statements. Once again, suicide can make sense to a client who has lost all hope.

The W in IS PATH WARM represents withdrawing and counselors should be concerned if a client begins to pull away from family and friends. For some people who are contemplating suicide, it is appealing to withdraw from family and friends as it minimizes the more painful interactions and decreases the possibility of considering the impact of the suicide on family and friends. The A stands for anger and counselor should be particularly attuned to clients whose affect changes and they become increasingly more hostile. Some clients feel like they are being forced into ending their own lives and are angry about their situation. In addition, one of the risk factors for suicide is the loss of a relationship (e.g., a divorce); hence, anger and revenge accompanied by loss of a relationship can be particularly lethal for some individuals. Some clients will become more reckless as they feel they have little to lose since they will soon be ending their lives; hence, the R stands for recklessness. The final letter in the mnemonic is M, which stands for mood change. These mood changes are dramatic and can even be positive. Some clients will experience a lift in negative affect as they have found a way to end their pain (i.e., suicide).

If there are indications that the client may be suicidal, the next step for the counselor is to assess if there is an actual suicide plan. The research indicates that counselors should be particularly concerned if clients have a plan for attempting suicide as those with a plan are more likely to make an attempt as compared to those who do not have a plan. Furthermore, the counselor should assess for the detail of the plan and the ability to carry it out. For example, a counselor should be more concerned if a client has already purchased a gun as compared to have a vague idea that he or she may take some pills. It should be noted, however, that impulsive clients who may not yet have developed a plan may still commit suicide if their situation deteriorates.

Granello (2010) advised counselors to be aware of both risk factors and warning signs. Sometimes it is difficult to determine the risk factors from the warning signs, but risk factors are often associated with demographic factors. In general, men are more likely than women to die from suicide, almost four times more likely (American Foundation for Suicide Prevention, 2014), although some studies show women are more likely to attempt suicide. Moreover, older Americans are disproportionately likely to die by suicide. In terms of race or ethnicity, white and Native American are at greater risk of attempting suicide as compared to other ethnic groups. The highest suicide rate is for white men over the age of 85.

Adolescent suicide has increased in the past 20 years; suicide is now the third leading cause of death among individuals 15 to 24 years old. Although suicide is somewhat rare for children before puberty, it does exist. For children ages 5 to 14, suicide is the seventh leading cause of death. Individuals responsible for children under the age of 18 have also been found to be at lower risk than individuals without those responsibilities.

In determining the risk of suicide, Sanchez (2001) also suggested that the counselor assess whether there are some "protective factors," which have been found to reduce the risk of attempting suicide. For example, being married or in a significant relationship seems to serve as a protective factor. In addition, people who have children under the age of 18 are less likely to attempt suicide. Having a general purpose for living and a support system also seem to be important protective factors. Other positive factors to consider are whether the individual is employed or involved in a structured program (e.g., going to school).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Counselors need to be proficient in detecting suicidal thoughts and assessing the intensity of those thoughts. Counselors must assess this information accurately to determine if clients are in immediate danger of harming themselves. If the client is deemed to be in immediate danger, then the practitioner must have procedures for responding quickly to this potentially deadly situation. Granello (2010) recommended erring on the side of caution and that it is better to explore suicide with someone who is not suicidal than to ignore possible signs with someone who is suicidal. Furthermore, Granello (2010) expelled the myth that talking with someone about suicide will cause him or her to consider suicide. The reality is that discussing suicide does not elevate the risk.

The following is an example of an initial session that explores the possibility of suicidal ideation. In this case, Mary is a 39-year-old woman whose husband has recently moved out of their home. Mary works as an accountant for a large accounting firm. Her husband told her three weeks before that he had met with a lawyer to initiate divorce proceedings and that he wants to have shared custody of their 11-year-old son, Sean.

Counselor: Can you tell me what brought you in to counseling?

Mary: My husband left me three weeks ago, and I'm having trouble coping with that. Although I had gotten angry with him for working so much, I never thought he would leave me. We've been together for 15 years, and the thought of not being together is just so difficult for me. It is so painful not to be with him.

Counselor: Tell me more about the pain.

Mary: It's almost like a physical pain. I feel empty and hurt and there are times that I can't stop crying. I can't describe the pain and the emptiness. I just hurt so much, and I can't find any way to stop the pain. I sometimes throw up because I'm in so much pain. I spend hours crying in the bathroom and I can't get myself to stop. I cannot imagine a life without him. I have begged him to come back but he tells me that he can't stay married to me. When I call him, he hangs up on me and says he can't take any more. I feel so helpless and I can't get him to listen. I never felt so terrible. . . . I can't think about anything else.

Counselor: You seem very hurt by his decision to leave?

Mary: Very hurt doesn't quite describe it. I have never felt so hurt and rejected. I feel like nothing matters except for Sean. I try to do things like cook dinner for Sean, but it's very hard. I'm so depressed and I can't find the energy to even cook dinner. The future looks so bleak . . . I don't want to think about living without him. I don't see a way to feel anything but pain. I wish I could find a way to end the emptiness and pain.

Counselor: Sometimes when people are in as much pain as you are in, they'll do anything to stop the pain, including committing suicide. Have you ever thought about suicide as a way to stop the pain?

Mary: Although I don't want to admit it, I have thought about killing myself because I think that's the only way I could stop this pain.

Counselor: I understand the desire to stop the pain. When you thought about stopping the pain, what have you thought about?

Mary: I just think about how nice it would be not to hurt so much and that feeling nothing would be so much better than how I feel now.

Counselor: When was the last time you felt like this?

Mary: Last night. I always try to cry in the bathroom so that Sean won't see how upset I am and I thought I'd like to climb in the shower and slit my wrist.

Counselor: Is that the first time you felt like killing yourself?

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Mary No, I guess I've thought about climbing in the shower a couple of times.
Counselor: It sounds like you've thought about killing yourself in the shower, but have you thought about other ways you might kill yourself?
Mary: Not really. It's during my spells when I can't stop crying that I think about this mess.
Counselor: Have you thought about how you'd cut your wrists?
Mary: I've thought that I would need to buy those old-fashioned razor blades.
Counselor: Have you bought those?
Mary No, I can't do that to Sean. I wouldn't want Sean to suffer, and it would hurt him.
Counselor: Mary, I know you are in a lot of pain and I worry that you might hurt yourself to try and stop that pain. Are there other ways than slitting your wrist that you have thought about to kill yourself?
Mary: No. I won't have a gun in the house and I don't have any pills. I am embarrassed that I told you about the shower.
Counselor: You don't need to be embarrassed because I understand the desire to stop the pain. I am, however, very worried about you and about you hurting yourself.
Mary: Because of Sean, I won't do anything. I don't want to saddle him with the pain of having his mother commit suicide. I just want the pain to abate for just a little while.
Counselor: I think we can work together and that the pain will get better, but that will take you coming to counseling. Right now I need some reassurances that you won't hurt yourself. I probably should inform the authorities to keep you from killing yourself. What kind of assurances can you give me that you won't hurt yourself?
Mary: To be honest, I can promise you that I won't kill myself.
Counselor: There are many things that we can do to help you and we need to start exploring those now. Let's start with ways you can take care of yourself and Sean and develop a strategy for those times you feel like climbing in the shower. I am also going to give you a phone number of a suicide hotline that I want you to call if you have an episode where you start to think about getting in the shower to hurt yourself. Also you need to promise me that you won't buy any razor blades or anything else that you might use to hurt yourself.
Mary: I can promise all of that.

Formal instruments have been constructed to assess the risk of suicide. Many clinicians, however, do not use these instruments in isolation; instead they use both interviewing and observation in conjunction with a formal instrument. It should be remembered that a counselor does not assess suicide risk just once. A suicide assessment is time-limited and pertains only to the risks at that time and under those current conditions. Therefore, counselors need to continually reevaluate the suicide potential of clients because circumstances often change and the risk of suicide can increase or decrease depending on changes within the client's life.

Suicide Potential Instruments. The *Suicide Probability Scale* (Cull & Gill, 1992) is a 36-item instrument that provides an overall measure of suicide risk. The scale includes subscales that assess hopelessness, suicidal ideation, negative self-evaluation, and hostility. An advantage of this instrument is that it includes separate norms for normal individuals, psychiatric patients, and lethal suicide attempters (those who have attempted suicide using a method that could have been lethal).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Aaron Beck, the noted cognitive therapist, has constructed two scales related to suicide potential. The *Beck Scale for Suicide Ideation* (Beck & Steer, 1991) can be used to identify and screen clients who will acknowledge suicidal thoughts. This scale uses five screening items that reduce the amount of time spent completing the scale for clients who are not reporting any suicidal ideation. The manual recommends that this instrument not be used alone but in combination with other assessment tools. Another of Beck's instruments, the *Beck Hopelessness Scale* (Beck & Steer, 1993), relates to measures of suicidal intent and ideation. This instrument also has norms available for suicidal clients. Another noted contributor in the area of assessment is William Reynolds (1988), who developed the *Suicidal Ideation Questionnaire*, which has two forms, one for grades 10 through 12 and a second (SIQ-JR) for grades 7 through 9. Reynolds (1991) also developed the *Adult Suicidal Ideation Questionnaire*.

Validating instruments that assess the potential to commit suicide is somewhat problematic. It would not be ethical to give a sample of potentially suicidal individuals an instrument and then simply sit back and collect data on which ones actually committed suicide. Hence, a counselor using one of these instruments needs to examine the validation information to determine what the instrument measures (e.g., suicidal ideation, correlation with depression measures, attempted suicide). None of the instruments designed to specifically identify suicide potential is sufficiently refined to use in isolation. To screen for suicide, the counselor needs both astute observational and interviewing skills, combined with trained clinical decision making. Some practitioners will also use other instruments—such as measures of depression, personality, or symptom checklists—in the global assessment of suicide potential. A counselor should consult with another knowledgeable professional if there are any concerns related to the assessment of suicide risk. Often with a suicidal client, a counselor may need to ensure that the client receives a psychiatric evaluation.

LO5 Assessment of Depression

Because of depression's direct connection to suicide, it is important for counselors to begin assessing for depression early in the counseling process. Because of its frequency and detrimental effects on individuals, some mental health clinicians assess the level of depression with every client. Depression is characterized by feelings of sadness, ranging from the common "blues" to severe depression. For information on what constitutes a major depressive episode in the American Psychiatric Association's (2013) *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition* (DSM-5), see Chapter 15.

To properly assess whether a client is depressed, a counselor needs familiarity with the common characteristics of depressed clients. For children and adolescents, the symptom of depression may be irritability rather than sadness and lack of interest. With some clients, the symptoms of depression are easily recognized. With other clients, there may be "masked" depression, with the individual concealing the dysphoric mood or denying the depressed feelings. In cases of masked depression, practitioners need to detect the subtle signs of depression and explore the issue further if initial suspicions of depression are confirmed.

According to McNamara (1992), the symptoms of depression usually include a dysphoric mood, often accompanied by feelings of anxiety, guilt, and resentment. Depression is characterized by difficulties with concentration, decision making, and problem solving. Depressed clients tend to be pessimistic and have negative attitudes toward themselves, others, and the world. Behavioral and physical symptoms are also usually present, particularly if the depression is more serious. Social withdrawal is another symptom, and depressed individuals are often prone to crying, lethargy, and loss of interest in activities that were previously pleasurable. In terms of physical symptoms, fatigue is almost always

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

present, sometimes accompanied by a loss of appetite and an inability to sleep. The reverse may also be true, with appetite and amount of sleep required substantially increased. Headaches, muscle aches, and decreased libido are other indicators of possible depression. In some cases, there may be a reduction in speech and a slowing of body movements. Included in the clinician's assessment of depression is whether the client should be referred for a psychiatric evaluation and possible medication.

The assessment of depression includes determining both the severity of the depression and the type of depression. Chapter 15, which addresses diagnosis, discusses different types of depression. Practitioners must be skilled in differentiating among the types of depression and assessing the risk of suicide. Sometimes clinicians will employ one of the following standardized instruments to assess the severity of the depression.

Beck Depression Inventory-II. Ponterotto, Pace, and Kavan (1989) identified 73 different measures of depression that researchers or mental health practitioners use. Of these different measures, the *Beck Depression Inventory* (BDI) was the most commonly used. In fact, the BDI was used 10 times more often than the second most popular instrument. Moreover, Peterson, Lomas, Neukrug, and Bonner (2014) found the BDI to be the most frequently used assessment by counselors. The latest version of this instrument is the *Beck Depression Inventory-II* (BDI-II; Beck, Steer, & Brown, 1996). The BDI-II was developed to align with the *Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition* (DSM-IV) criteria for depression. Like its predecessor, the BDI-II contains only 21 items and usually takes clients about five minutes to complete. With each item, there are four choices that increase in level of severity. Therefore, the BDI-II results provide an index of the severity of depression. With college students, Steer and Clark (1997) found an overall coefficient alpha for the BDI-II of .92 and support for the construct validity of the instrument. Dozois, Dobson, and Ahnberg (1998) found that the factor structure of the BDI-II is stronger than that of the BDI. Furthermore, Beck published the *Beck Youth Inventories*, one of which is the *Beck Depression Inventory for Youth*.

Other Depression Instruments. There are other standardized instruments that could also be used in the assessment of depression. For children and adolescents, there are the *Children's Depression Inventory-2* (Kovacs, 2010); the *Children's Depression Rating Scale-Revised* (Poznanski & Mokros, 1996); the *Reynolds Adolescent Depression Scale-2nd Edition* (Reynolds, 2002); and the *Reynolds Child Depression Scale-2nd edition* (Reynolds, 1989). For adults, another instrument (besides the BDI-II) is the *Hamilton Depression Inventory* (Reynolds & Kobak, 1995). The *Revised Hamilton Rating Scale for Depression* (Warren, 1994) also assesses the client's level of depression; however, the clinician rather than the client completes this instrument. In conclusion, a substantial number of depression assessments have been developed, so the practitioner has some options in selecting an assessment of depression.

LO6 Assessment of Potential Violence

In the initial interview, it is also important for counselors to assess clients for potential violence. Although child abuse would come under this broader term of violence, I will talk about it separately. With the advent of school shootings and shootings at movie theaters, there is significant interest in identifying individuals who are prone to violence. It is, however, difficult to assess for violence.

Deisinger, Randazzo, O'Neill, and Savage (2008) contended that the pathway to violence is a four-step process that involves the steps of ideation, planning, acquisition, and implementation. A counselor's involvement could be at any step in the process. For example, a counselor could be working with a client who has violent ideation and the next step would

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

be to ask them if he or she has a plan. If there is a plan, the next step would be to see if the client has acquired the weapons to implement the violence. If the client has already acquired the resources and has a plan, then the counselor should seek immediate consultation as confidentiality may need to be breached.

The American Counseling Association's (2014) *Code of Ethics* clearly state that an exception to confidentiality is when there is serious and foreseeable harm. Therefore, if a counselor believes that a client is going to cause harm to another then confidentiality may need to be broken. According to Van Brunt (2015), counselors should be familiar with the specifics of **duty to warn** in their state. Clients should be made aware of counselors' duty to warn as a part of the informed consent procedure. As a result of the *Tarasoff* case, counselors may need to inform an identifiable victim of his or her potential risk. In the case of a potentially violent client, a counselor should seek professional consultation immediately about issues related to duty to warn.

This discussion is not designed to provide an overview of threat assessment, which involves determining whether there is in actuality and real and potential threat (Van Brunt, 2015). Rather, this discussion focuses on signs of potential violence and how counselor should assess for violence. In terms of violence assessment, the old adage applies: The best predictor of future behavior is past behavior. Therefore, it is important in the initial interview to ask about previous violent behaviors.

In assessing clients for dangerousness, Van Brunt (2015) recommends that counselors cover several key factors. First, a counselor should assess the immediacy of the potential attack. More concern is merited when an attack is eminent as compared to planning an attack that may occur several months in the future. Van Brunt also encourages counselors to assess the degree to which the client is focused or fixated on the attack. The closer the attack and the more focused an individual becomes on a particular person, place, or system, the more likely the attack. Van Brunt further recommends assessing the level of organization or disorganization of the plan. Counselor should also consider legacy tokens, which are writings or media content prepared by the perpetrator prior to the attack that are designed to be found after the attack. Social isolation and access to weapons are other potential signs. There may also be a tendency by the perpetrator to depersonalize and objectify the target of the violence.

Violent behavior is very difficult to predict and there are not guaranteed methods for accurately assessing it. Counselors, however, should explore the topic of violence because they may be working with a victim or a potential perpetrator. In order to assess for violence, the topic of violence must be brought up.

L07 Assessment of Child Abuse

All 50 states have laws regarding child abuse and in most of these laws, counselors are specifically named as mandated reporters (Henderson, 2013). It is important to remember that counselors do not have to prove that there is abuse, but they must report to Child Protective Services (CPS) if they *suspect* that there is child abuse. According to the U.S. Department of Health and Human Services (2015), the definition of child abuse and neglect is:

> *Any recent act or failure to act on the part of a parent or caretaker which results in death, serious physical or emotional harm, sexual abuse or exploitation; or an act or failure to act, which presents an imminent risk of serious harm. (p. viii)*

Child abuse or child maltreatment usually falls into four categories: physical abuse, emotional abuse, sexual abuse, and neglect.

This discussion of assessing for child abuse is designed for the generic school or mental health counselor who is working with children and adolescents. This overview is not

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

intended for those investigating child abuse or professionals who may be interviewing children for legal recourse or for forensic purposes. Interviewing children or adolescents related to child abuse takes specialized training, which is beyond the scope of this book. The focus here will be on signs of child abuse and what to do if a child discloses that he or she has been abused.

The presence of a single sign does not mean that child abuse is occurring; however, a closer look at the situation may be warranted when signs appear repeatedly or in combination (Child Welfare Information Gateway, 2013). Counselors should consider the possibility of physical abuse if the child has unexplained bruises, burns, broken bones, or bites. Other signs include shrinking at the approach of adults or protest/cries when it is time to go home. Another sign is the child has a history of abusing animals or pets. Counselors might also consider physical abuse if the parent or other caregiver has a history of physically abusing the child or who uses harsh physical discipline. Another sign is when the parent or caregiver provides unconvincing or conflicting explanation about the child's injury or offers an explanation that is not consistent with the injury. A third sign is when the parent or caregiver describes the child in very negative terms, such as saying that the child is evil (Child Welfare Information Gateway, 2013).

Signs of emotional abuse are when the child exhibits extremes in behavior, such as significant passivity or aggression. The child may also act inappropriately like an adult or regress and act inappropriately infantile. Another sign of emotion maltreatment is a delay in either physical or emotional development. Counselor may also want to consider emotional abuse if the child has attempted suicide or if a child reports a lack of attachment to the parent. Signs that the parent or caregiver may be emotionally abusing the child are constantly blaming, belittling, or berating the child. Other signs include rejecting the child or being unconcerned about the child (Child Welfare Information Gateway, 2013).

A prominent sign of sexual abuse is if the child reports being sexually abused by a parent or another adult caregiver. Other signs of sexual abuse are if the child has difficulty sitting or walking, or reports nightmares or bedwetting. Another pertinent sign is if the child has sophisticated or unusual sexual behavior or knowledge. Another sign is suddenly refusing to change for gym or to participate in physical activities. Research has also found that children who have been sexually abused may attach very quickly to new adults in their environment or may run away from home. Furthermore, if the child becomes pregnant or contact a venereal disease, then the counselor should consider sexual abuse, particularly if the child is under age 14. Signs of parent or caregiver behaviors that may indicate sexual abuse are if the parent is unduly protective of the child or severely limits the child's contact with other children, particularly children of the opposite sex. Other signs include being very secretive, isolated, or jealous (Child Welfare Information Gateway, 2013).

Signs that the child may be neglected are the child frequently misses school or is consistently dirty and has severe body odor. Another sign of neglect is begging or stealing food or money. The child may also lack sufficient clothing for the weather and lacks needed medical or dental care. Unsupervised children may also abuse alcohol or other drugs. The child may also state that there is no one at home to provide care and supervision. Counselors should also consider neglect when the parent or other adult caregiver appears indifferent to the child or seems apathetic towards the child. Sometimes children will be neglected when parents are severely depressed or when they are abusing substances (Child Welfare Information Gateway, 2013).

Specific procedures for mandated reporting vary from state to state and counselors should be aware of the procedures in their state. In addition, most schools have a set procedure for reporting child abuse and school counselors must be aware of those procedures. Informed consent information should include a clause that confidentiality will be broken if child abuse is suspected.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

If a counselor suspects child abuse, then the first step, if possible, is to seek consultation from a senior clinician. Many times counselors can call personnel in agencies related to CPS and they will help the clinicians in determining if reporting is the best avenue available. Clinicians should remember that it is best to err on the side of being safe. On the other hand, clinicians should understand that children are susceptible and that questions may be leading, and it is better to leave the investigation to professional who know what they are doing. Therefore, counselors should use skills such as reflection and paraphrasing when a child is disclosing abuse (Henderson, 2013). It should be remembered that the counselor is only responsible for reporting suspected child abuse and the counselor's role does not extend to investigating or proving that there is abuse. According to Henderson (2013), when abuse is suspected, the primary objective is to listen and create a safe space for the child. From a forensic standpoint, counselors do not want to inadvertently taint or change the child's testimony so it is important for a clinician to be careful in their questions and approach.

LO8 Assessment of Substance Abuse

According to Adesso, Cisler, Larus, and Haynes (2004), the assessment of substance use is a relevant task for clinicians regardless of the population with whom or the setting in which the therapist works. On the other hand, Greene and Banken (1995) contended that clinicians frequently fail to recognize clients' symptoms as reflecting substance abuse or dependence. In 2009, an estimated 21.8 million Americans aged 12 or older were illicit drug users, which represented 8.7% of the population aged 12 and older. Illicit drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type drugs used nonmedically. Furthermore, in 2009, 6.8% of the population aged 12 or older or 17.1 million people were classified as being heavy drinkers. Among young adults aged 18 to 25, 41.7% reported binge drinking, and the rate of heavy drinking was 13.7% (Substance Abuse and Mental Health Services Administration, 2010). In the United States, Grant et al. (2004) found the highest prevalence of alcohol dependence among 18- to 20-year-olds who began drinking in early adolescents. Hence, substance use in the United States presents a serious problem and one that counselors need to consider in initial assessment.

Adesso et al. (2004) cited studies that indicated that in most mental health settings between 29% and 50% of the individuals seeking services will also have a substance use disorder. Problematic use of a substance(s) is often not a unitary disorder.

Given the number of clients who have a substance-related issue, it is important that counselors assess substance use early in the counseling process. A counselor who does not explore the possibility of substance abuse could easily proceed with addressing only issues that are essentially a result of the client's substance abuse. One method for exploring the possibility of substance abuse is simply to ask the client. Oetting and Beauvais (1990) found that self-reports were reasonably reliable. There are, however, some problems with validity, because some clients are reluctant to disclose the use of certain substances or the amount of use. Nevertheless, counselors should explore with every client what drugs they are taking and the amount of alcohol they typically consume. In terms of medications, use of prescription, over-the-counter, and street drugs should all be explored for several reasons. First, heavy use of any of these types of drugs can be problematic. A second reason for exploring all types of drugs is that even some commonly used over-the-counter medications have side effects. Sometimes a client's symptoms (e.g., being tired all the time) result from the medication; therefore, counselors need to consult the *Physician's Desk Reference* in order to be knowledgeable about the side effects of any drug the client is taking. Table 8.5 includes some examples of questions to ask clients, particularly if they indicate that they are drinking or taking drugs.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 8.5
Examples of questions related to substance use

1. Do you take any medications, over-the-counter drugs, or other drugs such as street drugs? If medication or over-the-counter drugs are used, for what and how much? If street drug, what and how much?
2. Do you drink alcohol? If yes, how many drinks in a typical week?
3. How often do you get drunk or high in a typical week?
4. When did you start drinking or taking the drug(s)?
5. Has anyone ever mentioned that this is a problem? Who? Why do you think they see it as a problem?
6. Where and with whom do you usually drink or use drugs?
7. Has drinking or taking drugs ever caused you any problems (e.g., problems with family, employers, friends, or the law)?
8. Have you ever tried to quit? What happened? Were there any withdrawal symptoms?
9. Is your drinking or drug use causing you any financial or other types of difficulties?
10. Do you ever mix drugs or mix drugs and alcohol? Do you ever substitute one substance with another substance (e.g., do you ever use amphetamines instead of cocaine)?

In assessing substance use, counselors need to focus on assessing which substances are used and the amount taken. Other areas that should be included in assessment of substance abuse are the social and interpersonal aspects because this information will be helpful in the treatment phase. Treatment will vary depending on whether the client is drinking or using alone or with friends. Another goal is to identify the internal and external triggers that precede the use of the substance. For counselors working with adolescents, it is important that clinicians understand the risk and protective factors related to substance abuse. Counselors interested in this area are directed to the National Institute on Drug Abuse website (http://www.drugabuse.gov/publications/preventing-drug-abuse-among-children-adolescents/chapter-1-risk-factors-protective-factors/what-are-risk-factors).

Motivational Interviewing. Let me start by saying motivational interviewing (MI) is more than an assessment strategy, it is a treatment approach that stresses a relational component that focuses on empathy and a technical component that involves the differential responses that reinforce clients' change talk (Miller & Rose, 2009). Furthermore, although motivational interviewing began with individuals with substance use disorders, it is now used to address various aspects of behavioral change. In this discussion, however, I will use substance use as an example of motivational interviewing. In contrast to the more confrontational assessment, where the goal is to have the individual admit their problem with substance use, Miller and Rollnick (2002) advised therapists to respond with a empathic person-centered style with special attention on evoking and strengthening the client's own motivations for change. Motivational interviewing is based on four general principles: express empathy, develop discrepancy, roll with resistance, and support self-efficacy.

Expressing empathy is a core tenet of the theory and is central to any type of MI intervention/assessment. The second tenet or principle is developing discrepancy, which is based on the concept that clients have ambivalence about their substance use (i.e., wanting to change but at the same time not wanting to alter their behavior). The counselor purposefully does not focus on "sustain talk" so not to continue to elicit and strengthen it but, rather, intentionally focuses on change talk. The focusing on change talk will increase the discrepancy between current use and wanting to change. The third principle is rolling with resistance. As indicated earlier, MI is not a confrontational style but one that focuses on empathy and change; therefore, if resistance is encountered then the counselor either

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

focuses on empathizing with the resistance or focusing on the language of change. The final principle concerns supporting self-efficacy. Here, the counselor conveys to the client that he or she wholeheartedly believes in the client's abilities to make changes (Glover, 2010). There is considerable empirical support for MI (see Miller & Rose, 2009). Motivational interviewing, however, is not a technique or a simplistic method of counseling; it is a complex and challenging treatment approach.

Standardized Instruments in Substance Abuse. A number of standardized instruments are designed specifically to assess substance abuse. One instrument is the *Substance Abuse Subtle Screening Inventory* 3 (SASSI-3; Miller, 1997). The SASSI Institute purported that the SASSI-3 has a 93% rate of accuracy in identifying individuals with substance-related disorders. There is also the SASSI-A2 (Miller, 2001), a screening instrument for adolescents who may have a substance use disorder.

Another assessment tool that is often used in medical settings is the CAGE, which is an interviewing technique. The CAGE (Mayfield, McLeod, & Hall, 1974) is a mnemonic device for the following four questions:

1. Have you ever felt you need to *cut down* on your drinking?
2. Have people *annoyed* you by criticizing your drinking?
3. Have you ever felt bad or *guilty* about drinking?
4. Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (*eye opener*)?

If the answers to these questions indicate a problem, then the clinician should gather more detailed information about the client's drinking.

Sometimes, however, clients will deny that drinking or drug use is a problem. One technique to help clients admit a drinking or drug use problem is to have them self-monitor and record all drinking and drug use. Some clients will be surprised by the amount recorded. Another technique, sometimes referred to as the "acid test," is to ask clients to control or limit their drinking. For example, a client is asked to have no more than three drinks at any one occasion for three months. This technique provides an indication of whether the drinking can be controlled.

In the assessment of substance abuse, counselors must also assess whether they can provide the appropriate services for a client. Some clients will need to be referred to specialists in the area of substance abuse or to an outpatient or inpatient facility specializing in substance abuse.

L09 Mental Status Examination

Although not a formal psychometric instrument, counselors in many hospitals and mental health settings need to be able to conduct and write the results of a **mental status examination** (MSE; Polanski & Hinkle, 2000). The mental status examination is used to describe a client's level of functioning and self-presentation and is generally conducted during the initial session or intake interview. The MSE is not necessarily a separate and distinct portion of the interview, although it is often integrated into the initial counseling process. Morrison (2007) concluded that the mental status examination is simply a statement of how a person looks, feels, and behaves at the moment of the initial examination. The information gathered in an MSE is often used for diagnostic purposes and to organize client information. The MSE is frequently organized around the following categories: (1) appearance, attitude, and activity; (2) mood and affect; (3) speech and language; (4) thought process, thought content, and perception; (5) cognition; and (6) insight and judgment (Trzepacz & Baker, 1993).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Although there are some variations among the precise categories of information included in a mental status examination, most procedures start with the clinician noting the client's general *appearance*, including such physical attributes as grooming, dress, cleanliness, and any notable abnormalities. Polanski and Hinkle (2000) suggested that counselors also include a notation on *attitude* and physical activity in this first category. In terms of attitude, counselors should note a client's presenting attitude and approach to the interview, including such descriptors as *cooperative, hostile, guarded, tense,* or *regressed*. Because the level and type of physical activity can indicate neurological problems, substance abuse, anxiety, or hyperactivity, Polanski and Hinkle also suggested including documentation on the level and quality of a client's physical or motor movements. A mental status examination report may also include pertinent physical movements or manifestations (e.g., crying, laughing, pacing, grunting) that are either inappropriate or incongruent with the context.

A client's *mood and affect* is typically the second category on which a clinician remarks in an MSE. Morrison (2007) suggested that clinicians consider three qualities when discussing mood: its type, lability, and appropriateness. According to Trzepacz and Baker (1993), most moods can be partially captured by the descriptors of euthymic (normal), dysphoric, euphoric, angry, anxious, and apathetic. When reporting on mood, it is often useful for counselors to include the client's own words. Lability concerns the degree to which moods seem to change in a given time frame. Some clients present with a high degree of lability and their moods change frequently during an initial interview, whereas other clients exhibit a more consistent mood through the first interview. In addition, some moods are appropriate (e.g., depression after the death of a spouse) and other clients may exhibit moods that may not coincide with their circumstances. In a mental status examination, a clinician usually describes the client's mood in a few sentences.

As the third category, Polanski and Hinkle (2000) recommended that clinicians include analysis of *speech and language* in a mental status examination. Morrison (2007) labeled this section "flow of speech." In particular, a counselor should note derailments in speech, loose associations, flights of ideas, tangential speech, poverty of speech, or pressured speech. As will be discussed in Chapter 15, characteristics of the client's speech can be very helpful in determining a diagnosis.

The fourth major category included in a mental status examination includes *thought process*, *thought content*, and *perception*. Analysis of thought process and content not only is important in diagnosing psychotic disorders but can also be useful in diagnosing mania, depression, dementia, drug intoxications, and some personality disorders. Once again, counselors should note unusual thought processes or content in an MSE. As with mood, the clinician should consider whether the thought content is consistent with the client's current circumstances (Morrison, 2007).

Cognition is typically the fifth category assessed in the MSE. Cognition refers to intellectual functioning and the ability to think logically and rationally. Morrison (2007) suggested that clinicians also consider the intellectual resources that the client may be able to use in further treatment. Unlike other portions of the MSE, this section can frequently involve a formal instrument, with the most common being the Mini-Mental State Exam (Folstein, Folstein, & McHugh, 1975). The *Mini-Mental State Examination*, 2nd Edition, has recently been published with a new normative group of more than 1,500.

Finally, Polanski and Hinkle (2000) indicated that counselors should include an analysis of *insight and judgment*. In this case, insight refers to clients' awareness of their own personality and behaviors that may be contributing to their current difficulties.

In managed care situations and other mental health situations, clinicians are often required to write an MSE report or an abbreviated summary of the client's presenting issues and background. Polanski and Hinkle (2000) suggested that this report include pertinent psychosocial background information in addition to the six categories listed previously.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The focus of the report should be on describing the client's behavior during the initial session or intake procedure, and the order of the content is not typically prescribed. Often in managed care situations, the MSE report must include information that supports the diagnosis. (Detailed information on diagnosis is covered in Chapter 15.) The following is an example of an MSE report.

> *Michael is a 33-year-old Caucasian male employed as a research analyst for a department within his state government. Michael came to counseling because his supervisor has insisted that he seek help because he has missed a number of workdays in the past three months and is substantially behind on a number of projects. Michael appeared appropriately dressed for work in a professional office, but his clothes were wrinkled, and he indicated he had not had time to shower in the past three days. Michael reports that his supervisor believes Michael's problems at work may be related to his drinking and occasional drug use. Michael reports that he sometimes drinks to excess on weekends and snorts cocaine about once a week. His speech seemed somewhat pressured, and he often skipped from topic to topic and had difficulty staying focused on many of the topics. He also complained of "cotton mouth." Michael appeared to be of above-average intelligence and seemed logical and rational in his cognitions. No gross or fine motor impairments were noted, but there appeared to be some occasional tremors in both hands. He reports he is under a physician's care but is not taking any prescribed medications. Although he expressed some "anxiety" about the possibility of losing his job, his mood often appeared euphoric. He consistently laughed, smiled, and appeared jovial, even when discussing difficult issues. He seemed to avoid questions related to the amount of his drinking and drug use and "couldn't remember" if he had used cocaine in the past week. Insight into the degree to which substance use may play into his difficulties seemed limited.*

Summary

The initial information-gathering process in counseling frequently involves the counselor interviewing the client. The skill with which the counselor performs this initial assessment has an influence on the overall effectiveness of the counseling. Because interviewing is one of the most frequently used counseling assessment tools, counselors should consider the reliability and validity of their own interviews. Within the initial, or intake, interview, it is important that counselors not only identify the most pressing problems but also explore these issues from diverse perspectives. Gathering detailed information about the presenting concerns provides a sound foundation for the selection of interventions and treatment approaches. In gathering client information, structured, semistructured, and unstructured interviews each have strengths and limitations, and the choice of which to use depends on the setting and the primary purposes of the clinical interviews.

Some counselors incorporate formal and informal assessment measures into the initial paperwork completed by the client. Checklists and rating scales can be efficient methods for gathering information. Commonly used instruments are the Symptom Checklist-90—Revised or the shorter version of this instrument, the Brief Symptoms Inventory. These measures of symptoms can also be used throughout the counseling process to monitor changes in symptoms (e.g., feeling anxious) as a result of the counseling. With some of the checklists and rating scales for children, a parent or guardian completes the forms rather than the child. In addition to using standardized instruments, counselors can develop their own informal checklists and rating scales; however, these should be used with extreme caution.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The counselor's initial assessment of a client also needs to consider specific issues, such as potential suicide, depression, violence, child abuse, or substance abuse. Assessment of these potential problem areas should involve a multifaceted approach. Because of the lethal aspects of suicide, counselors should routinely include an investigation of suicidality in their counseling. An assessment of suicidal risk should include an investigation of possible intent, whether there are plans and a specific method, the progress made in instituting the plan, and an evaluation of the client's mental state, particularly the degree of hopelessness. The evaluation of suicide potential should not be a single step but rather a continual monitoring and evaluating of the potential risk.

Another area that is important to assess is whether the client has been a victim of violence or is a potential perpetrator. Just as with suicide, this is not a one-time event and counselors should be sensitive to signs of violence throughout the counseling process. Although assessment of violence is very difficult, the best predictor of future violent behavior is previous violent behaviors.

Counselors also need to consider child abuse in their initial assessment of children. Child abuse can be physical, emotional, or sexual and the signs vary depending on the type of abuse. Specific procedures for mandated reporting vary from state to state and counselors should be aware of the procedures in their state. Counselors should remember that they do not have to prove that child abuse has occurred, but they must report to child protective services if they suspect child abuse.

Because depression is a common disorder, counselors need to be aware of the subtle signs of depression so that they can assess the severity of the depression. Substance use is another area counselors need to include in their initial assessment. Counselors should systematically explore the use of both drugs and alcohol with clients because this is a prevalent problem. The assessment of substance abuse needs to include inquiry into the amount of the substance(s) used, whether the use can be controlled, and the consequences of the use.

In some settings, a counselor may be required to write a summary of a mental status examination. A mental status examination is typically not a "formalized" assessment, although some settings may have a structured interview and then the clinician writes the mental status examination after completing the interview. As part of conducting a mental status examination, practitioners sometimes will briefly assess the client's level of intellectual functioning by using the *Mini-Mental Status Exam* or the recently released *Mini-Mental Status Exam*, 2nd Edition.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 9

Intelligence and General Ability Testing

Learning Objectives

- **LO1** Describe major models or theories of intelligence
- **LO2** Understand the advantages of individual intelligence tests
- **LO3** Identify the major individual intelligence tests
- **LO4** Distinguish among the primary index scores of the *Wechsler Intelligence Scale for Children—Fifth Edition* (WISC-V)
- **LO5** Understand the advantages and disadvantages of group intelligence tests
- **LO6** Describe the major issues in intelligence testing

As a school counselor in an elementary school, you receive a voice-mail message from the mother of Justin, a fourth grader at your school. Justin has recently been tested for the school's academically accelerated, or "gifted," program, and his mother is upset because Justin was determined not eligible for the program. In your school district, a student must have an intelligence quotient (IQ) of 130 to be in this accelerated program. Justin's mother wants information on the intelligence test used, and she questions the test's legitimacy because she says she knows her son and is convinced that he is highly intelligent. Justin's mother is quite upset that her son was not accepted into the accelerated program with an IQ score of 123, and his mother questions whether 7 points really reflect a significant difference in intellectual capabilities. As you prepare to return this parent's phone call, you begin to investigate the intelligence test used by the school psychologist in your district. Furthermore, you begin to think about what it means to be highly intelligent and how we actually assess intellectual abilities. In addition, you begin to contemplate what an IQ score of 130 actually represents and if it is significantly different from a score of 123.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

As indicated in the discussion of Justin, intelligence or general ability tests often are used to determine if a student is placed in an accelerated program in a school. In addition, these same assessments can be used in making decisions about the most appropriate instructional strategies for students with cognitive difficulties. Very few counselors will be involved in the administration of individual intelligence tests in their careers; however, many counselors will be involved in activities that require the knowledge of intelligence testing. A school counselor may be part of a committee to determine if a child qualifies for special education. Mental health counselors frequently work with individuals who have certain cognitive impairments that may have been diagnosed with an intelligence test. A rehabilitation counselor may work with a client who has recently experienced a traumatic brain injury. In all of these cases, knowledge of intelligence testing will assist counselors in better serving their clients.

Although terms such as *intelligence* and *IQ* are a common part of everyday language, there is a lack of consensus within the profession concerning the specifics on definition and structure of intelligence. In the United States, Sternberg, Conway, Ketron, and Bernstein (1981) found that both laypeople and experts had some common ideas about intelligence that included some kind of problem solving, a factor related to verbal ability, and some kind of social competence. Some professionals in the field of assessment prefer the term *general ability* because of the negative connotations associated with intelligence testing. The debate about the meaning of intelligence and intelligence test scores is somewhat emotionally charged. According to Neisser et al. (1996), political agendas, rather than scientific knowledge, often fuel the debate related to intelligence testing. To facilitate an understanding of intelligence testing, this chapter briefly reviews several different perspectives on the nature of intelligence. This overview of models lays a foundation for the discussion of specific intelligence or general ability tests. Both individually administered and group tests are included in the discussion. This chapter concludes with a summary of research related to pertinent issues in intelligence or general ability testing.

LO1 Models of Intelligence

Psychometric Approach

Since the beginning of the twentieth century, a major influence on the study of individual differences in intelligence has been the *psychometric* (or differential) *approach*. Psychometric theories of intelligence are based on the premise that intelligence can be described in terms of *mental factors* (Bjorklund, 2012), which are general mental skills that influence mental performance in a variety of situations. The term *factor* is used in this approach because many theorists have based their assertions on the statistical technique of factor analysis. From this theoretical perspective, researchers need to identify one or more factors of intelligence and then develop intelligence tests to measure differences on those identified factors. Theorists, however, are not in total agreement on the number or types of factors that constitute intelligence.

At one extreme in terms of the number of factors in intelligence is **Spearman's model** (1927). Spearman was one of the first theorists to discuss factors. He postulated a two-factor theory of intelligence. He contended that everyone has the first factor—a *general ability factor* (*g*) that influences a person's performance on all intellectual tasks. Spearman contended that the second type of factors was specific factors that influence performance in specific areas. According to Spearman, these specific factors were all, to some degree, correlated with *g*, and that in essence intelligence is a rather homogeneous construct. To put it very simply, smart people are smart, and unintelligent people are not smart. This homogeneous view of intelligence has had considerable influence on people's views of intelligence.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

At the other end of the continuum concerning the number of factors in intelligence is Guilford's (1988) structure-of-intelligence theory, which includes 180 unique intellectual factors organized around three dimensions. Figure 9.1 illustrates this complex model in which the first dimension is *mental operations* and contains six forms. These mental operations can involve five *content areas*, which is the second dimension. Finally, there are six possible *products*, the third dimension, that interact with the combinations of operations and content areas.

Between the extremes of Spearman and Guilford is another early theorist, Thurstone, who also had considerable influence on others' views of intelligence. Thurstone (1938) originally proposed a model of seven *primary mental abilities* rather than one solitary factor. In several studies, Thurstone demonstrated that seven common factors collectively accounted for most of the individual differences associated with intelligence. These seven primary mental abilities were:

1. Verbal comprehension
2. Word fluency
3. Number facility
4. Perceptual speed
5. Memory
6. Space
7. Reasoning

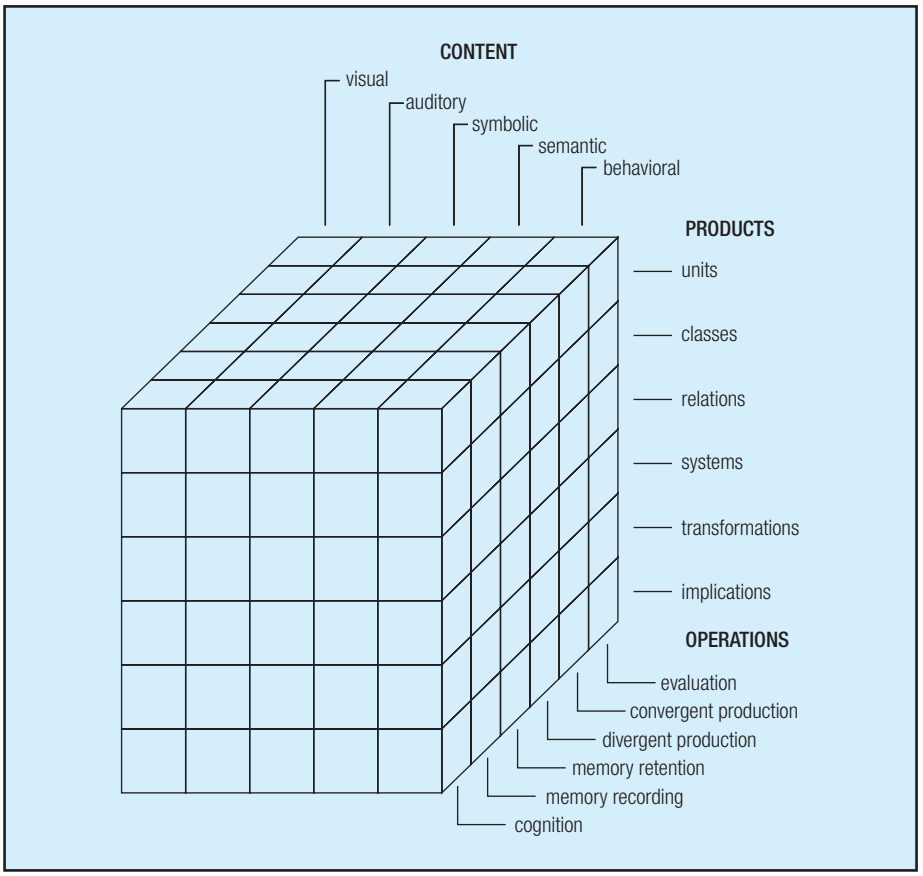


FIGURE 9.1
Gutford's structure-of-intelligence model. From *Educational and Psychological Measurement*, Vol. 48, pp. 1–4. Reprinted by permission of Sage Publications

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Thurstone's original theory did not include a general ability factor (*g*), but he did propose that the different factors were somewhat related to each other.

Other factor theorists primarily used a hierarchical approach. For example, Vernon (1950) contended that at the top of the hierarchy was *g* that influenced all other factors. According to Vernon, there were two second-order factors: verbal and educational aptitudes (v:ed) and spatial, mechanical, and practical aptitudes (k:m). Finally, other abilities branched down from these levels to these two second-order factors into third, fourth, and other order factors.

Cattell-Horn-Carroll Model. Although some consider the Cattell-Horn-Carroll model of cognitive abilities to be one of the more influential contemporary theories (Evans, Floyd, McGrew, & Leforgee, 2002), this is not a theory that three theoreticians collaborated on, but rather a theory that has evolved from the work of three individuals. Cattell (1971, 1979) was a theorist who suggested that *g* was composed of two second-order factors. In Cattell's model, the second-order factors are *fluid abilities* (Gf) and *crystallized abilities* (Gc). Cattell proposed that fluid abilities are primarily biologically determined and affect all fields of relational perception. Because fluid abilities are considered relatively culture-free, they are often reflected in memory span and spatial thinking tests. Crystallized abilities, on the other hand, include acquired skills and knowledge that are influenced more by cultural, social, and educational experiences. Tests of verbal comprehension and social relations tap crystallized abilities. Horn and Cattell (1966) and Horn (1991) expanded Cattell's Gf-Gc model by adding several other intelligences (i.e., short-term acquisition and retrieval, visual intelligence, auditory intelligence, long-term storage and retrieval, cognitive processing speed, correct decision speed, quantitative knowledge, and reading and writing skills). Based on factor analysis and other statistical techniques, Carroll (1997) concluded that there was convincing findings that supported a general factor of intelligence. Furthermore, Carroll (1997) used an analogy based on geology in which intelligence is a hierarchical structure. The top stratum is *g* or general intelligence; the second stratum that underlies the first stratum is made up of eight abilities and processes; and the third stratum below each of the second stratum abilities is comprised of more specific factors that relate to and support the eight second stratum factors. McGrew (1997) proposed that the similarity between the Cattell-Horn model and the Carroll model reflected a single model, which he labeled the Cattell-Horn-Carroll Theory of Cognitive Abilities.

More similarities exist among the psychometric approaches to intelligence than might at first appear. Toward the end of their careers, Spearman and Thurstone became more similar in their views. Spearman acknowledged that separate group factors (e.g., verbal abilities, mathematical reasoning) exist, and Thurstone concluded that correlation exists among his primary factors, suggesting the possibility of one general factor. Overall, most psychometricians agree that there is both one general ability factor (*g*) *plus* a number of lower-level factors that reflect more specific skills. Differences among the theorists primarily relate to the extent to which intelligence can be described by *g* and the content of the lower-level factors (Bjorklund, 2012).

Developmental Progressions

Rather than focusing solely on the structure of intelligence, developmental theorists have suggested that intelligence can be better understood by examining how intelligence develops. The most well-known developmental theorist is probably Jean Piaget (1972). Unlike the psychometricians, Piaget had little interest in how individuals might differ in intellectual abilities because he was interested in children's development. Others have adapted Piaget's concepts to the analysis of individual differences

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

and to measuring intelligence. One of the ways in which Piagetian approaches have influenced the measuring of intellectual functioning is through the examination of individual differences in development.

Piaget suggested that intelligence involves a developmental progression, with individuals moving sequentially through the stages of the *sensorimotor*, *preoperational*, *concrete*, and *formal operations* periods. Children move through these successively higher stages of intellectual development through the use of two intellectual functions. One of these is *assimilation*, which is the process by which a child incorporates a stimulus into an existing cognitive structure or schema. The second process is *accommodation*, a process by which the child creates a new schema or significantly alters an existing schema based on new information. According to Piaget, cognitive development does not occur only through maturation; learning and the environment also influence the process. Piaget's tasks can be modified to serve as measures of individual differences; however, the tasks have not directly influenced the major standardized tests of intelligence.

Another developmental theorist whose work is having an increasing influence on the field of education is the Russian psychologist Vygotsky. His theories are complex and currently are having a significant influence on instruction but, currently, little influence on intelligence testing.

Some researchers have recently been interested in biological development, particularly the study of the brain as the basis for measuring intelligence (Neisser et al., 1996). With the increasing sophistication in brain imaging technology, it soon may be possible to understand more about the influence of biological development and specific characteristics of brain functioning.

Related to biology is Ceci's (1990, 1993) bioecological theory, which is a theory that also incorporates a developmental perspective in the conceptualization of intelligence. Ceci believes that biological factors have a major influence on intelligence; however, he contends that there is no single underlying factor (*g*), nor is human intelligence highly heritable. Ceci, along with others, sees intelligence as being multifaceted, with individuals varying in the domain-general abilities. These theorists posit that there are no intelligent people; there are only people who are intelligent in certain domains. *Context* is central to Ceci's theory, for he argued that intellectual abilities are highly influenced by the context in which they are performed. For example, one of his studies showed that children learned better when the context was a video game versus a laboratory. The context in which people learn affects the knowledge base they possess, an idea that is central to Ceci's theory. Thus, Ceci argued, an individual's intelligence might not be reflected in the methods currently used to assess intelligence. He further contended that intelligence is reflected in knowledge acquisition and in the ability to process, integrate, and elaborate on the knowledge in a complex manner.

Information Processing

Rather than studying the structures or development of intelligence, information-processing models focus on how individuals process information. Hence, the focus is not on *what* we process (factors of intelligence) but on *how* we process. One information-processing approach that has had an influence on intelligence testing is the work of Luria (1966). Luria proposed that intellect could best be described by two methods of processing information: simultaneous processing and sequential processing. **Simultaneous processing** involves the mental ability to integrate input all at once. To solve simultaneous processing items correctly, an individual must mentally integrate fragments of information to "see" the whole. **Sequential processing** involves a different set of processing skills that requires the individual to arrange stimuli in sequential or serial order to solve a problem. The Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004) was influenced by Luria's theory.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

A theory of intelligence influenced by Luria is Naglieri and Das's (1997, 2005) Planning, Attention, Simultaneous, and Successive (PASS). Naglieri and Das adopted Luria's theorization that certain areas of the brain are associated with basic psychological process. These areas of the brain facilitate four interrelated cognitive processes that are at the heart of intelligence: Planning, Attention, Simultaneous, and Successive process. *Planning* is a frontal lobe function and is associated with strategies needed to complete tasks for which a solution is needed. It involves self-monitoring, impulse control, evaluation, and generation. *Attention* is a mental process that is closely associated with orienting response. The longer the tasks requires attention, the more difficult the process. Simultaneous processing essentially involves organizing information into groups or a meaningful whole. With simultaneous processing, the individual is able to see patterns as interrelated elements. On the other hand, successive processing involves arranging the stimuli in a specific serial order. Successive processing is needed when sounds or movements are required in a certain order, such as learning successive sounds in early reading. According to Naglieri and Das (2005), intelligence tests should focus on these processing abilities rather than the traditional approach that is more similar to achievement tests. The Cognitive Assessment System (CAS; Naglieri & Das, 1997) is an assessment developed that operationalizes the PASS theory.

Another information-processing theory that has received some attention is Sternberg's (1985, 1988) triarchic theory. This theory of intelligence is quite extensive and incorporates three subtheories: (1) the internal world of the individual or the mental processes that underlie intelligence, (2) the experiential subtheory, and (3) the individual's contextual or external world. The information-processing portion of the theory is part of the first subtheory. According to Sternberg, an individual's ability to process information depends on the use of three components: the metacomponents, the performance components, and the knowledge-acquisition components. The *metacomponents* are the higher-order executive processes that are used to plan what to do, monitor it while it is being done, and then evaluate it after it is completed. Consider, for example, the interaction between planning abilities and intellectual capabilities. Sternberg argued that more intelligent people focus on critical issues that need to be solved, whereas less intelligent people attend to less important situations. The metacomponents influence the selected *performance components* (e.g., encoding, mental comparisons, and retrieval of information). Once again, individuals vary in their abilities to use these performance skills. According to Sternberg, intelligence is related to selecting the appropriate strategies for processing diverse information. The third component, *knowledge acquisition*, involves individuals selectively encoding, combining, and comparing information to add to their knowledge base. Thus, Sternberg argued, differences in intelligence are a reflection of differences in people's abilities to process information at the metacomponent, performance, and knowledge-acquisition levels.

The second subtheory within Sternberg's triarchic theory concerns the relationship between experience and intelligence. This subtheory relates to an individual's ability to process information within existing experience. According to Sternberg, intelligence is best measured when the task is rather novel or when the task is in the process of becoming automatized. In Sternberg's view, the ability to deal with novelty is a good measure of intelligence. In essence, people who are more intelligent are able to react more quickly and process new information more efficiently than people who are less intelligent. Another of Sternberg's contentions is that the ability to automatize is also an indicator of intellect. If a person is effective at automatizing information, then that individual will have more resources available for dealing with novelty.

The third subtheory within Sternberg's theory concerns the contextual or external world of the individual. Sternberg argued that intelligence is contextually bound. For example, having an uncanny knowledge of elephants' migratory patterns is not highly valued in many settings in Western society. Thus, people will vary in their abilities to select,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

adapt to, and shape environments, which are also, according to Sternberg, marks of intelligence. Although Sternberg's theory has not yet had a direct influence on intelligence instruments, his contention that intelligence tests only measure a selective portion of the complex construct of intelligence is having an influence on the understanding of what intelligence is. Most of us probably know someone who would not do particularly well on the instruments discussed later in this chapter, even though that person may be "street smart" or "have an uncanny ability to size up the situation." Sternberg is a leader in promoting the expansion of methods to view and measure intelligence.

Other Theories

Another theorist who advocated for expanding our views of intelligence is Gardner (1993). Gardner's theory of multiple intelligences avers that any set of adult competencies that is valued in a culture merits consideration as a potential intelligence. As the name of his theory implies, he argued that there are *multiple* intelligences rather than a single entity. Gardner's theory is different from others, such as Thurstone, in that he proposes nine relatively independent frames of mind: (1) linguistic, (2) logical-mathematical, (3) musical, (4) spatial, (5) bodily kinesthetic, (6) interpersonal, (7) intrapersonal, (8) naturalist, and (9) existential. Gardner suggested that current psychometric tests only measure linguistic, logical, and some aspects of spatial intelligence, while ignoring the other intelligences. He argued that measures need to value intellectual capacities in a wide range of domains and that the methods should be appropriate for each domain (e.g., for spatial ability, testing an individual's ability to navigate in a new city). There are some difficulties with adapting Gardner's ideas in assessment, but his theory has attracted considerable attention in the educational assessment area.

In conclusion, the different theoretical approaches have influenced the assessment of intelligence. The psychometric approach has had the most direct influence on the instruments commonly in use today. The newer theories, however, provide some important insight into the complexities of intelligence and may influence future methods of assessment. Defining and conceptualizing intelligence is often perplexing, which means the measuring of intelligence is also plagued by many of these same issues.

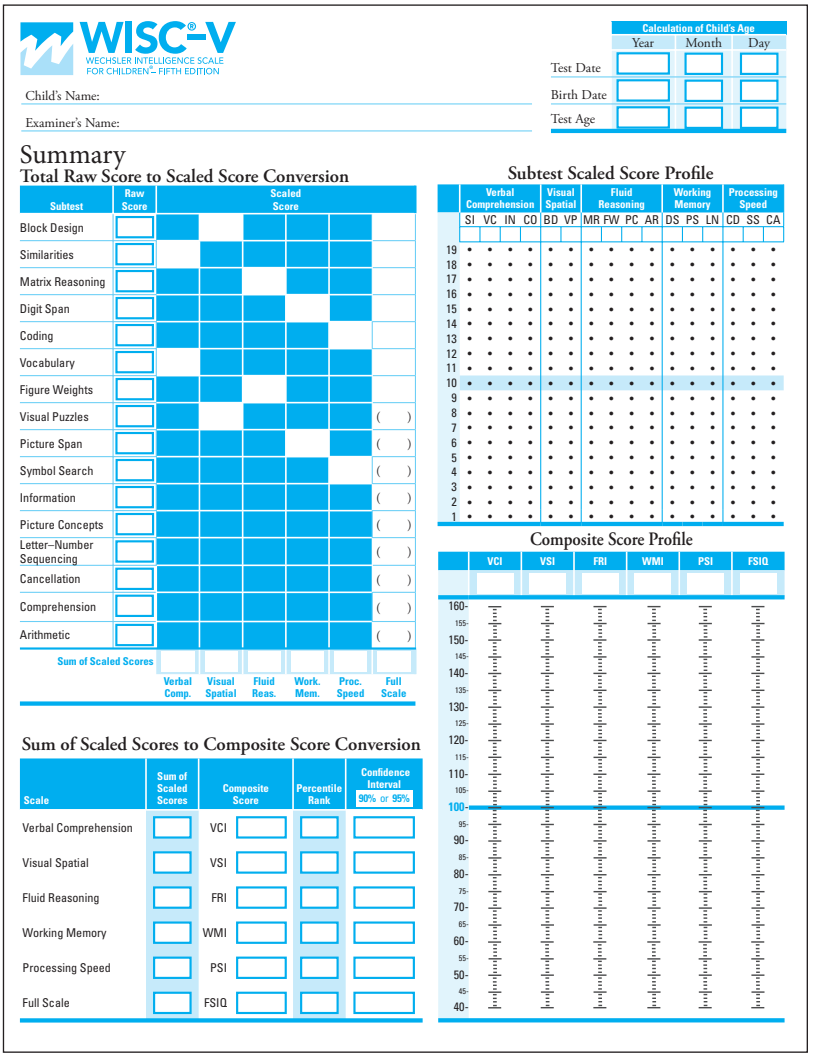
L02 Individual Intelligence Testing

As stated earlier, counselors are typically not responsible for the administration, scoring, and interpretation of individual intelligence tests. In order to administer an individual intelligence test, a professional must have completed certain coursework and received specific training and supervision on the use of that instrument. This chapter does not provide the necessary information to become proficient in the administration and interpretation of these instruments. Rather, the intent of this chapter is to give an overview of commonly used tests so that counselors can use the results in counseling contexts or when participating in multidisciplinary teams involving intelligence assessment. This overview includes a brief description of the most widely used instruments.

L03 Wechsler Scales

David Wechsler had extraordinary influence on the testing of intelligence. Wechsler (1944) defined intelligence as "the capacity to act purposefully, to think rationally, and to deal effectively with his [or her] environment" (p. 3). He developed a series of three intelligence instruments that are similar in approach but designed for different age groups. The *Wechsler Preschool and Primary Scale of Intelligence IV* (WPPSI-IV; Wechsler, 2012) is designed for children ages 2 years, 6 months through 7 years, 7 months old; the *Wechsler Intelligence Scale*

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.



WISC®-V
WECHSLER INTELLIGENCE SCALE FOR CHILDREN®–FIFTH EDITION

Child's Name:

Examiner's Name:

Calculation of Child's Age: Year | Month | Day

Test Date

Birth Date

Test Age

Summary

Total Raw Score to Scaled Score Conversion

Subtest	Raw Score	Scaled Score
Block Design		
Similarities		
Matrix Reasoning		
Digit Span		
Coding		
Vocabulary		
Figure Weights		
Visual Puzzles		( )
Picture Span		( )
Symbol Search		( )
Information		( )
Picture Concepts		( )
Letter–Number Sequencing		( )
Cancellation		( )
Comprehension		( )
Arithmetic		( )
Sum of Scaled Scores		
	Verbal Comp.	Visual Spatial	Fluid Reas.	Work. Mem.	Proc. Speed	Full Scale

Subtest Scaled Score Profile

	Verbal Comprehension				Visual Spatial		Fluid Reasoning				Working Memory			Processing Speed		
	SI	VC	IN	CO	BD	VP	MR	FW	PC	AR	DS	PS	LN	CD	SS	CA
19																
18																
17																
16																
15																
14																
13																
12																
11																
10																
9																
8																
7																
6																
5																
4																
3																
2																
1																

Composite Score Profile

VCI	VSI	FRI	WMI	PSI	FSIQ

160, 155, 150, 145, 140, 135, 130, 125, 120, 115, 110, 105, 100, 95, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40

Sum of Scaled Scores to Composite Score Conversion

Scale	Sum of Scaled Scores		Composite Score	Percentile Rank	Confidence Interval 90% or 95%
Verbal Comprehension		VCI			
Visual Spatial		VSI			
Fluid Reasoning		FRI			
Working Memory		WMI			
Processing Speed		PSI			
Full Scale		FSIQ			

FIGURE 9.2
Summary Form for the *Wechsler Intelligence Scale for Children—Fifth Edition*. Copyright © 2014 by NCS Pearson, Inc. Reproduced with permission. All rights reserved. "Wechsler Intelligence Scale for Children" and "WISC" are trademarks, in the United States and/or other countries, of Pearson Education, Inc. or its affiliate(s).

for Children—Fifth Edition (WISC-V; Wechsler, Raiford, & Holdnack, 2014) is for children ages 6 years, 0 months through 16 years, 11 months old; and the *Wechsler Adult Intelligence Scale—Fourth Edition* (WAIS-IV; Wechsler, 2008) is for individuals 16 through 90 years, 11 months old. In addition to these three major assessments, there are also the *Wechsler Abbreviated Scale of Intelligence* (WASI) and the *Wechsler Intelligence Scale for Children—Fourth Edition Integrated* (WISC-IV Integrated). The Wechsler scales are commonly cited in research, and numerous reports on test usage indicate that the Wechsler instruments are the most frequently used intelligence tests (Archer & Newsom, 2000). Furthermore, even if we examine all types of tests (e.g., personality), Peterson, Lomas, Neukrug, and Bonner (2014) found that the Wechsler Intelligence Scale for Children was the seventh most frequently used instrument by counselors. Therefore, counselors should be familiar with these instruments and with their strengths and limitations.

The earlier versions of the WPPSI-IV, the WISC-V, and the WAIS-IV included measures of Full-Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ). This, however, appears to be changing as the most recent revisions of these assessments have a measure of FSIQ and a number of index scores. With the WISC-V, there are five index

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

scores: Verbal Comprehension Index (VCI), the Visual Spatial Index, the Fluid Reasoning Index (FRI), the Working Memory Index (WMI), and the Processing Speed Index (PSI). With the WPPSI-IV, WISC-V, and WAIS-IV, the FSIQ is the global and aggregate measure of cognitive abilities and has a mean of 100 and a standard deviation of 15. The other indexes or composite scores also have a mean of 100 and a standard deviation of 15. Due to the similarities of all the Wechsler intelligence instruments, the following information on the *Wechsler Intelligence Scale for Children—Fifth Edition* (WISC-V; Wechsler, Raiford, & Holdnack, 2014) should also acquaint readers with many aspects of the WPPSI-IV and the WAIS-IV.

The revision of the WISC was based on various structural models of intelligence and cognitive ability, neurodevelopmental and neurocognitive research, psychometric results, clinical utility, and the needs of practitioners. The WISC-V can be administered by only someone trained in its administration. The trained clinician has two choices: the traditional paper-and-pencil format and the new digital format available on Q-interactive. Q-interactive uses two iPads that are connected by Bluetooth where the examiner has one iPad and the child being tested has the other. Q-interactive also includes other instruments, such as the WIAT-III or the KTEA-3 (see Chapter 10), which the WISC-V was designed to coincide with. Significant research went into equating the results of these two administrative choices (Wechsler et al., 2014).

The administration and scoring of the Wechsler subscales are somewhat complex because different points are awarded depending on the quality of the responses. Items in each subtest are arranged hierarchically, beginning with the easier items and ending with the more difficult items. Similar to the Stanford-Binet, on certain subtests, the administrator uses a reverse rule procedure in determining the *floor* or *basal level*, which involves moving to easier items if a certain number of items are missed at the beginning of the subtest and then proceeding through the testing until a *ceiling level* is reached (i.e., a certain number of items are consecutively missed). Throughout the administration of the scales, the examiner also gathers anecdotal information, such as task-approach skills, level of concentration, and motivation. A qualified examiner can extract rich and important information from the performance on these intelligence tests as well as from observing the individual during the testing.

The WISC-V is composed of 21 subtests; however, a child would very rarely take all 21 subtests. There are various levels of interpretation of the WISC-V including the Full-Scale level, the Primary Index level, the Ancillary Index scale level, and the Complementary Index scale level. These four methods of interpretation require different subtests to be given, and different composite scores are utilized depending on the interpretation method. As this is a basic introduction to the WISC-V, this chapter will cover only the Full-Scale level and the Primary Index level.

LO4 Psychologists will often give 16 subtests in order to score the WISC-V at the Full-Scale level (see Table 9.1). The subtests listed in italics in Table 9.1 represent the subtests that are used to calculate FSIQ. With this approach, a full-scale IQ (FSIQ) is calculated and five factor-based primary index scores are determined. The five primary index scores are the Verbal Comprehension Index, the Visual Spatial Index, the Fluid Reasoning Index, the Working Memory Index, and the Processing Speed Index. With the Primary Index level, only 10 subtests are given (see Table 9.2) and the focus is solely on the five primary index scores mentioned in the previous sentence. The *Verbal Comprehensive Index* measures verbal reasoning with an emphasis on work knowledge acquisition, information retrieval, verbal problem solving, and general knowledge. The *Visual Spatial Index* assesses spatial reasoning and the ability to analyze visual details. The *Fluid Reasoning Index* measures abstract conceptual reasoning and the ability to conceptualize information abstractly and to apply that knowledge. The *Working*

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 9.1
Common subtests for the WISC-V

Verbal Comprehension	Visual Spatial	Fluid Reasoning	Working Memory	Processing Speed
■ Similarities	■ Block Design	■ Matrix Reasoning	■ Digit Span	■ Coding
■ Vocabulary	■ *Visual Puzzles*	■ Figure Weights	■ *Picture Scan*	■ *Symbol Search*
■ *Information*		■ *Picture Concept*	■ *Letter-Numbering Sequence*	■ *Cancellation*
■ *Comprehension*		■ *Arithmetic*		

TABLE 9.2
Subtest used to calculate five primary index scores

Verbal Comprehension	Visual Spatial	Fluid Reasoning	Working Memory	Processing Speed
■ Similarities	■ Block Design	■ Matrix Reasoning	■ Digit Span	■ Coding
■ Vocabulary	■ Visual Puzzles	■ Figure Weights	■ Picture Scan	■ Symbol Search

Memory Index assesses a person's ability to identify visual and auditory information, maintain it in temporary storage, and resequence it for use in problem solving. The *Processing Speed Index* measures the ability to rapidly identify visual information, make accurate decisions quickly, and rapidly implement those decisions.

The following describes each of the subtests that contribute to the FSIQ and the five primary index scores.

Verbal Comprehension Subtests

- *Similarities.* For the Similarities subtest, the child is read two words that are connected and the child need to identify the connection. The Similarities subtest is designed to measure verbal concept formation and abstract reasoning. It also measures crystallized intelligence, word knowledge, categorical thinking, and verbal expression.
- *Vocabulary.* With the Vocabulary subtest, the child is presented with pictures and verbal items. For the pictures, the child simply names the object and for the verbal items, the child defines the word that is read aloud. Vocabulary is designed to measure word knowledge and verbal concept formation.
- *Information.* Although Information is used to calculate FSIQ, it is also considered a Verbal Comprehension subtest. With Information, the child answers questions that are designed to be general knowledge topics. It measures the child's ability to acquire, retain, and retrieve information. It is also a measure of crystallized intelligence and long-term memory.
- *Comprehension.* Besides contributing to FSIQ, Comprehension is considered a measure of Verbal Comprehension. With Comprehension, the child answers questions based on an understanding of general principles and social situation. Comprehension is a measure of verbal reasoning and verbal comprehension and expression and the ability to use past experience and demonstrate practical knowledge and judgment.

Visual Spatial Subtests

- *Block Design.* For Block Design, the child views a model and/or picture that can be made with the blocks provided. The child is timed to see how long it takes him or her to complete the design. It is primarily a measure of the ability to analyze and

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

synthesize abstract visual stimuli. It also measures nonverbal concept formulation and reasoning, broad visual intelligence, simultaneous processing, and the ability to separate visual stimuli.

- *Visual Puzzles*. Visual Puzzles is used in calculating both FSIQ and the Visual Spatial Index score. With Visual Puzzles, the child views a completed puzzle and then selects three response options that, when combined, reconstruct the puzzle. The subtest measures mental rotation, visual working memory, understanding part-whole relationships, and the ability to analyze and synthesize abstract visual stimuli.

Fluid Reasoning Subtests

- *Matrix Reasoning.* With Matrix Reasoning, the child views an incomplete series or matrix and selects the response option that completes the matrix or series. This subtest requires the child to use visual spatial information to identify the underlying conceptual rule that links the stimuli and then apply the underlying concept to select a response. The subtest is designed to measure fluid intelligence, broad visual intelligence, classification and spatial ability, and simultaneous processing.
- *Figure Weights.* Figure Weights is a new subtest, and, within a specified time period, the child views a scale with missing weight(s) and then selects the response option that keeps the scale balanced. This task requires the child to apply the quantitative concept of equality. This subtest measures quantitative fluid reasoning and induction.
- *Picture Concept.* Picture Concept contributes to FSIQ, but it is also considered a Fluid Reasoning subtest. For Picture Concept, the child views two or three rows of pictures and selects one picture from each row to form a group with a common characteristic. The test measures the child's ability to identify the underlying conceptual relationship among objects. The subtest measures fluid and inductive reasoning, visual-perceptual recognition and processing, and conceptual reasoning.
- *Arithmetic.* The final subtest that contributes to FSIQ is Arithmetic, which is also considered a measure of Fluid Reasoning. For Arithmetic, the child mentally solves arithmetic problems within a specified time. Arithmetic involves mental manipulation working memory, numerical reasoning ability, applied computational ability, and mental alertness.

Working Memory Subtests

- *Digit Span.* In Digit Span, the child is read a series of numbers, which he or she must recall in the same order. Then the child is read a series of numbers and must reverse the order. After that, the child is read a series of numbers and then the child must say them from the smallest number to the largest. All of the Digit Span exercises require registration of information, brief focused attention, auditory discrimination, and auditory rehearsal.
- *Picture Span.* For Picture Span, the child views a page with one or more pictures for a specified period of time and the child selects the pictures from a different page. Second, the child must select the pictures in the same order as they are presented. Picture Span measures visual working memory and working memory capacity and also involves attention, visual processing, and response inhibition.
- *Letter-Numbering Sequencing.* While contributing to FSIQ, Letter-Numbering Sequencing is considered a Working Memory subtest. With this subtest, the child is read a series of numbers and letters and is required to recall the numbers in ascending order and then the letters in alphabetical order. It measures cognitive processing, auditory discrimination, attention, concentration, registration, and auditory rehearsal.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Processing Speed Subtests

- *Coding.* With Coding, it uses a key to copy symbols that correspond with simple geometric shapes or numbers. This is a timed subtest. The subtest measures processing speed, short-term visual memory, psychomotor speed, visual perception, visual scanning ability, attention, concentration, and motivation.
- *Symbol Search.* In Symbol Search, the child scans or searches a group of symbols to see if the target symbol(s) are there and indicates no if they are not present. The subtest measures visual-perceptual abilities and decision making. It also measures visual memory, visual-motor coordination, inhibitory control, sustained attention, and concentration.
- *Cancellation.* Cancellation contributes to FSIQ and is considered Processing Speed subtest. While working with a specified time limit, the child scans two arrangements of objects (one random and one structured) and marks target objects. The subtest measures the speed of visual-perceptual recognition and discrimination and also involves attention, concentration, and visual recall.

The interpretation of the WISC-V typically involves more than analyzing the results of the primary index scores and the subtest scores. The psychologist usually uses a process of identifying strengths and weaknesses of the five primary index scores and ten of the subtests; however, the Q-interactive system facilitates this process. The process also involves examining pairwise difference comparison of the primary index scores and certain subtest scores. There are also ancillary index scores that can be calculated and analyzed. There are also complementary subtests with accompanying complementary index scores (e.g., Naming Speed, Symbol Translation, and Storage and Retrieval).

The norming sample for the WISC-V included 2,200 children aged 6.0–16.11. A stratified sample was selected to match the demographics of the 2012 U.S. Bureau of Census regarding the variables of age, sex, race/ethnicity, parent education level, and geographic region. There are 11 age groups, and each of the age groups consisted of 200 individuals. There is extensive information on the reliability of the WISC-V in the technical and interpretative manual (Wechsler, Raiford, & Holdnack, 2014), which includes estimates of internal consistency, test-retest, and interscorer agreement. An internal consistency estimate for the full-scale IQ (FSIQ) is .96 with the following estimates for the primary index scores: .92 for the Verbal Comprehension Index, .92 for the Visual Spatial Index, .93 for the Fluid Reasoning Index, .92 for the Working Memory Index, and .88 for the Processing Speed Index. Standard error of measurement estimates are provided for all scores in the manual.

It is difficult to evaluate the validity evidence for the WISC-V as it is a new instrument. Confirmatory factor analyses tend to confirm the internal structure of it. The manual also provides evidence of the degree to which the WISC-V correlates with other measures of intelligence or cognitive ability, measures of achievement, and behavioral assessments. There are also special group studies that have been conducted with a variety of groups, including those who are intellectually gifted and those with moderately severe intellectual disabilities. Although the validity evidence is just beginning to be conducted on the WISC-V, this instrument builds on previous editions and there is a wealth of validity evidence for these previous editions.

Other Related Wechsler Instruments. An extension of the WISC-IV is the WISC-IV Integrated (Wechsler et al., 2004) that focuses on identifying processing problems and involves the administration of a selection of 16 process subtests in addition to the administration of the WISC-IV. There are plans for a WISC-V Integrated, but it has not been published at the time of writing this book. The WISC-IV Integrated is not an updated version of the WISC-IV, but it is designed to expand on the information gathered in the WISC-IV particularly as it relates to identifying information processing strengths and limitations. The

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

process subtests expand on the WISC-IV subtests, and a knowledgeable psychologist can provide more detailed information about a child's cognitive processes.

For many reasons including the baby boomers getting older, there is increased interest in adults' memory and neurological disorders. The *Wechsler Memory Scale IV* (WMS-IV; Wechsler, 2009) is an individually administered measure of memory and working memory that can be used in conjunction with the WAIS-IV or independently. It was designed to measure episodic memory, declarative memory functioning, and working memory and requires substantial training to use effectively (Chittooran, 2010). It also contains a brief measure of cognitive status. There are two WMS-IV batteries: the Adult battery is for individuals ages 16 to 69 and the Older Adult battery are for individuals 65–90. The index scores for the Adult version are Auditory Memory, Visual Memory, Visual Working Memory, Immediate Memory, and Delayed Memory. The Older Version takes less time to administer and has fewer subtests. The WMS-IV is co-normed with the WAIS-IV, which can expand the clinical usages of the WMS-IV.

In conclusion, the Wechsler instruments are the most widely used individually administered intelligence tests. A skilled psychologist can derive an abundance of useful information from the administration of the WPPSI-IV, WISC-V, or WAIS-IV. Although most interpretations of the results will focus on cognitive abilities, some psychologists are able to use these testing results to identify personality factors and learning styles. Spanish versions of the Wechsler scales are also available and have been researched and validated.

Stanford-Binet Intelligence Scale

One of the oldest and most widely used intelligence assessments is the *Stanford-Binet*, which is now in its fifth edition (Roid, 2003). The Stanford-Binet evolved from the Binet-Simon, which was first published in 1905 and contained only 30 items. The Binet-Simon was revised in both 1908 and 1911. The first version of the Stanford-Binet was published in 1916 under the leadership of Terman at Stanford University. This version of the instrument was the first to include an intelligence quotient (IQ), derived from the ratio of mental age to chronological age. Major revisions of the Stanford-Binet were performed in 1937, 1960, 1972, and 1986. The fourth edition abandoned the ratio IQ in favor of a deviation IQ. The fifth edition of the Stanford-Binet (SB5) was built on this extensive historical experience and the substantial research that has been conducted related to earlier versions of the Stanford-Binet. In addition, some of the procedures were revised in order to address criticisms of earlier editions, particularly related to measuring the nonverbal aspects of intelligence and to making the assessment more engaging for diverse groups.

The current version of the Stanford-Binet was constructed for use with individuals 2 through 85 years old. According to Becker (2003), the SB5 contains an equal balance of verbal and nonverbal content. Norm-referenced scores are determined for Verbal IQ (VIQ), Nonverbal IQ (NVIQ), and Full-Scale IQ (FSIQ). There are five factors for both the verbal and nonverbal areas. These are:

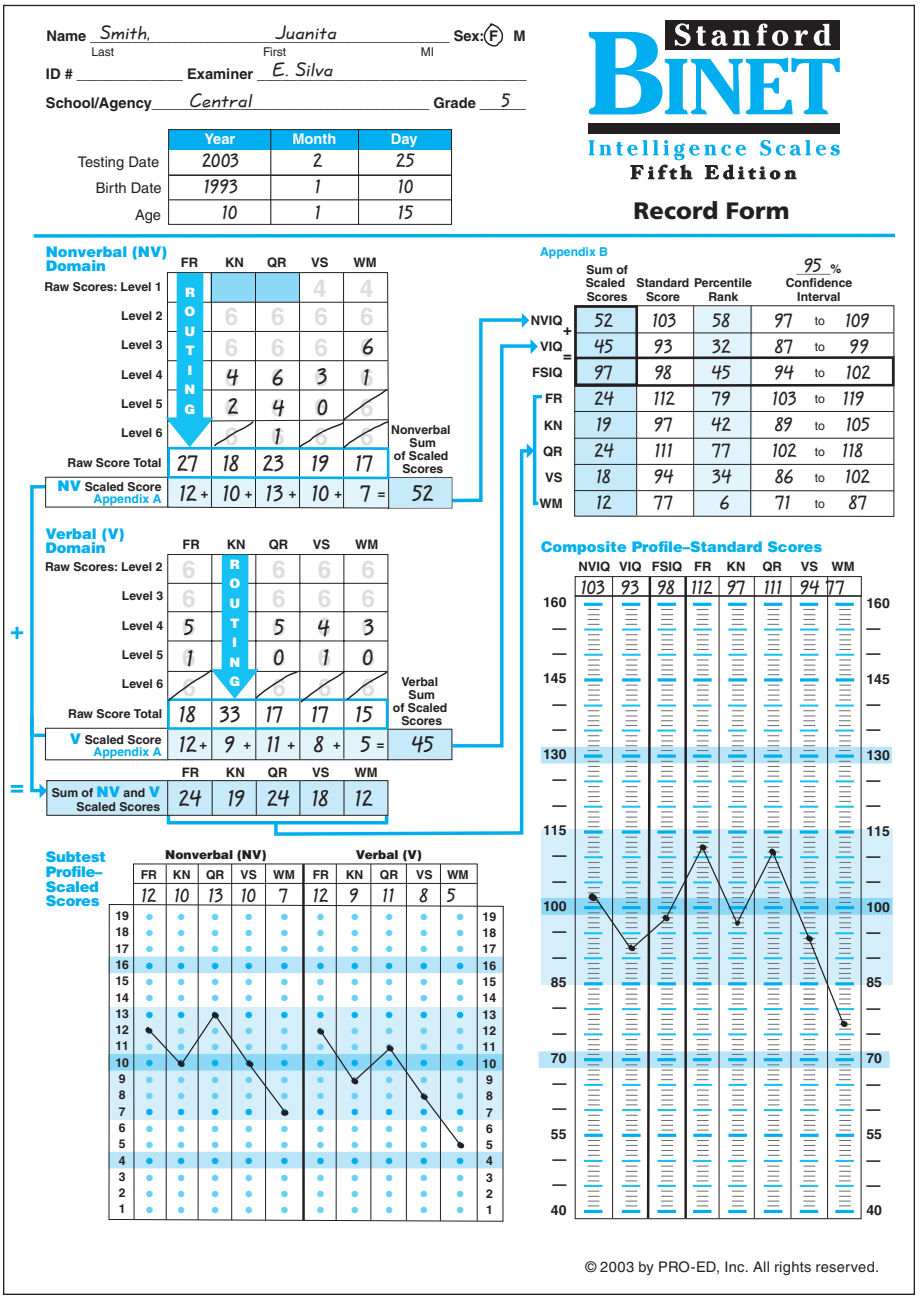
1. *Fluid Reasoning* concerns an examinee's abilities to solve verbal and nonverbal problems using inductive and deductive reasoning.
2. *Knowledge* assesses an examinee's accumulated fund of general information acquired at home, school, work, or through life experiences.
3. *Quantitative Reasoning* measures facility with numbers and numerical problem solving, whether with word problems or figural relationships (emphasis is on problem solving more than mathematical knowledge).
4. *Visual-Spatial Processing* involves an examinee's abilities to see patterns, relationships, spatial orientation, and the gestalt among diverse aspects of visual displays.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

5. *Working Memory* measures short-term memory processing of verbal and visual information, with an emphasis on "transformation" or "sorting out" of diverse information.

Similar to the Wechsler instruments, the Stanford-Binet's three intelligence score composites (i.e., FSIQ, VIQ, and NVIQ) have a mean of 100 and a standard deviation of 15 (see Figure 9.3). Previous editions of the Stanford-Binet had a mean of 100 and a standard deviation of 16. Within the Nonverbal Domain there are five Scaled Scores (e.g., Nonverbal Fluid Reasoning, Nonverbal Knowledge) and within the Verbal Domain there are five

FIGURE 9.3
Stanford-Binet Intelligence Scale, Fifth Edition. From the *Stanford-Binet Intelligence Scales, Fifth Edition (SB5) Examiner's Manual* (Figure 5.3, p. 152) by Gale H. Roid. Copyright © 2003 PRO-ED, Inc. Used with permission


Name Smith, Juanita Sex: F M
Last First MI
ID # Examiner E. Silva
School/Agency Central Grade 5
Year Month Day
Testing Date 2003 2 25
Birth Date 1993 1 10
Age 10 1 15
Stanford BINET
Intelligence Scales
Fifth Edition
Record Form
Nonverbal (NV) Domain
FR KN QR VS WM
Raw Scores: Level 1 4 4
Level 2 6 6 6 6
Level 3 6 6 6 6
Level 4 4 6 3 1
Level 5 2 4 0
Level 6 1
ROUTING
Raw Score Total 27 18 23 19 17
NV Scaled Score Appendix A 12 + 10 + 13 + 10 + 7 = 52
Nonverbal Sum of Scaled Scores
Verbal (V) Domain
Raw Scores: Level 2 6 6 6 6
Level 3 6 6 6 6
Level 4 5 5 4 3
Level 5 1 0 1 0
Level 6
ROUTING
Raw Score Total 18 33 17 17 15
V Scaled Score Appendix A 12 + 9 + 11 + 8 + 5 = 45
Verbal Sum of Scaled Scores
Sum of NV and V Scaled Scores 24 19 24 18 12
Appendix B
Sum of Scaled Scores Standard Score Percentile Rank 95 % Confidence Interval
NVIQ 52 103 58 97 to 109
VIQ 45 93 32 87 to 99
FSIQ 97 98 45 94 to 102
FR 24 112 79 103 to 119
KN 19 97 42 89 to 105
QR 24 111 77 102 to 118
VS 18 94 34 86 to 102
WM 12 77 6 71 to 87
Composite Profile–Standard Scores
NVIQ VIQ FSIQ FR KN QR VS WM
103 93 98 112 97 111 94 77
160 145 130 115 100 85 70 55 40
Subtest Profile–Scaled Scores
Nonverbal (NV) Verbal (V)
FR KN QR VS WM FR KN QR VS WM
12 10 13 10 7 12 9 11 8 5
19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
© 2003 by PRO-ED, Inc. All rights reserved.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Scaled Scores (e.g., Verbal Fluid Reasoning, Verbal Knowledge) that have a mean of 10 and a standard deviation of 3. The Scaled Scores for each domain are combined to produce five composite factor scores (i.e., FR, KN, AR, VS, and WM) that also have a mean of 100 and a standard deviation of 15 (see Figure 9.3).

The SB5 contains items and content from the fourth edition of the Stanford-Binet; however, the administration procedures are somewhat different. The administration of the SB5 starts with administering the Nonverbal Reasoning Routing Subtest (Object Series/Matrices) and then moving to the Verbal Knowledge Routing Subtest (Vocabulary). Once the routing tests have been administered, the administrator begins testing in the Nonverbal Domain until a ceiling has been reached for Nonverbal Knowledge, Nonverbal Quantitative Reasoning, Nonverbal Visual-Spatial Processing, and Nonverbal Working Memory. Once the ceiling levels are established in the nonverbal area, the administrator then begins to assess the Verbal Domain by completing the activities related to Verbal Fluid Reasoning, Verbal Quantitative Reasoning, Verbal Visual-Spatial Processing, and Verbal Working Memory.

The standardization sample for the current version of the Stanford-Binet was more than 4,800 individuals. The reliability estimates, using a measure of internal consistency for the Full-Scale IQ, is .98. The Stanford-Binet is a commonly used instrument that historically has been influential; however, it is not as commonly used as the Wechsler series.

Kaufman Instruments

Other prominent names in the field of intellectual assessment are Alan Kaufman and Nadeen Kaufman who are a husband-and-wife team of psychologists with extensive experience in test development. In the 1980s, they published the *Kaufman Assessment Battery for Children* (KABC; Kaufman & Kaufman, 1983a, 1983b), which reflected many advances in intelligence testing, but also was the focus of some controversy. They next published the *Kaufman Adolescent and Adult Intelligence Test* (KAIT; Kaufman & Kaufman, 1993). More recently, the *Kaufman Assessment Battery for Children, Second Edition* (KABC-II; Kaufman & Kaufman, 2004) was published, which is grounded in Luria's processing model and the Cattell-Horn-Carroll model. The Kaufman instruments are not as widely used as the Wechsler instruments and, therefore, are not covered in the same degree of detail.

The Kaufman Assessment Battery for Children, Second Edition, was designed to assess children ages 3 to 18. Both the KABC and the KABC-II were designed with the intention of fairly assessing children from diverse backgrounds and minimizing ethnic group differences in performance. The KABC-II yields four or five scales depending on whether the examiner is using the Cattell-Horn-Carroll (CHC) or the Luria approach. From a CHC perspective, the KABC-II yields measures of Short-Term Memory, Visual Processing, Long-term Storage and Retrieval, Fluid Reasoning, and Crystallized Abilities. A composite of these five scales is the Fluid-Crystallized Index. From a Luria perspective, the scales are named Sequential Processing, Simultaneous Processing, Learning Ability, and Planning Ability. The global score from the Luria approach is titled Mental Processing Index. Kaufman and Kaufman (2004) recommended that examiners use the CHC approach unless the examinee is likely to have deficiencies in crystallized abilities due to nonstandard learning experiences. Braden and Ouzts (2005), however, contended that the KABC-II does not measure intended constructs consistent with Luria's theory and that marketing it as utilizing both theories is misleading.

The KABC-II has 18 subtests; however, an examiner would not use all 18 with one child. The selection of subtests is based on the child's age and which of the theoretical approaches is used for scoring. The KABC-II was standardized with a sample of 3,025 and the makeup of the sample was reflective of the demographics of the 2001 U.S. Census. The reliability coefficients, particularly test-retest, vary somewhat depending on the theoretical approach,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

but for the major components from both approaches, the range is from .77 to .92. The validation evidence is complex, but both Braden and Ouzts (2005) and Thorndike (2005) question the degree to which the instrument measures Luria's theory.

Woodcock-Johnson IV

Another instrument that has recently been revised is the *Woodcock-Johnson IV* (Schrank, McGrew, Mather, & Woodcock, 2014). The Woodcock-Johnson IV is more than an intelligence test; in fact, it contains three separate co-normed batteries that measure cognitive abilities, achievement, and oral language. These three batteries are the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG), the Woodcock-Johnson IV Tests of Achievement (WJ IV ACH), and the new Woodcock-Johnson IV Tests of Oral Language (WJ IV OL). The batteries can be given independently, but they are often given together. There are many ways to use the Woodcock-Johnson IV, but counselors will often see it used in the diagnosis of learning disabilities. The interpretation of it can involve discrepancies in test scores and cluster scores. The publisher of the Woodcock-Johnson IV now offers web-based scoring and reporting with the purchase of the Woodcock-Johnson IV.

The WJ IV COG measures general and specific cognitive functions and is based on the evaluation of Cattell-Horn-Carroll model of cognitive abilities. There are 10 tests in the standard battery and another 8 in the extended battery. No examinee is given all the tests. One method of using the WJ IV COG is to get a measure of Gf-Gc Composite, which measures fluid and crystallized cognitive abilities. The Gf-Gc score can then be compared to other measures of cognitive abilities, achievement scores, and performance in the oral language domain. These comparisons are done to identify individual strengths and limitations.

The WJ IV ACH is an individually administered achievement measure, and it includes 11 tests in the standard battery and 10 tests in the extended battery. Also, because the WJ IV ACH may be given multiple times, there are three parallel forms (A, B, C). The tests are related to reading, writing, spelling, and math and are designed to assess specifics within the achievement areas. Once again, there are numerous methods for interpreting the results of the WJ IV ACH either alone or in combination with the measures of cognitive abilities and oral language. For example, an examiner may find that a student's reading abilities are significantly lower than his or her oral language abilities.

Completely new to the Woodcock-Johnson IV is the WJ IV OL, which measures oral language proficiencies. This battery can be used with the WJ IV COG and the WJ IV ACH for diagnostic purposes. It is also possible to measure oral language proficiency in Spanish that can be compared to language proficiency in English if there is someone who can administer the tests in Spanish. This can be particularly useful in making instructional decisions about students whose first language is not English.

As will be discussed in Chapter 10, the methods for diagnosing learning disabilities are changing. An older method involved examining differences between intellectual abilities and achievement, and the WJ IV could easily be used with this approach to diagnosing learning disabilities. The newer approach to diagnosing learning disabilities is called Response to Interventions (RTI), which involves assessing how students respond to academic interventions. In RTI, the assessment processes are drawn out as a student's responses to different interventions related to possible learning disabilities are assessed using a variety of assessment approaches. The Woodcock-Johnson IV can also be used with an RTI approach as the WJ IV ACH can be given multiple times to assess whether the interventions are having a positive influence on achievement. The strength of the Woodcock-Johnson IV is its comprehensiveness in measuring cognitive abilities, achievement, and oral language. It is primarily used in the diagnosis of learning disabilities.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Additional Individual Instruments

A few other individually administered intelligence tests are used sufficiently to merit a brief overview. The *Differential Ability Scales—Second Edition* (DAS-II; Elliot, 2007) is based on the British Ability Scales, which has been used in England since 1983. The DAS-II has been standardized in the United States. As the name implies, the focus of the DAS is on identifying specific abilities. It is designed to provide insight, particularly to school psychologists, into how a child processes information. The diagnostic subtests measure abilities such as verbal and visual memory, immediate and delayed recall, visual recognition and matching, processing and naming speed, phonological processing, and understanding numerical concepts.

Another frequently used instrument is the *Slosson Intelligence Test-Revised, Third Edition* (SIT-R3; Nicholson & Hibpshman, 2002). The SIT-R3 provides a relatively quick assessment of cognitive abilities for children and adults. It takes approximately 15 to 20 minutes to administer and score (Erford, Klein, & McNinch, 2007). It consists of 187 items that are given orally; the examinee's language skills will influence performance. The norms for the SIT-R3 were recalibrated in 1998. This shorter assessment is not quite as reliable as the longer, individually administered intelligence assessments. On the positive side, the SIT-R3 is highly correlated with the Verbal Scale and Full-Scale IQ scores of the older Wechsler instruments.

As discussed later in this chapter, there has been criticism of intelligence tests being culturally biased. In response to this criticism, some individually administered intelligence tests have been developed that focus on trying to diminish the effects of cultural biases and the influences of language proficiency on intelligence. For example, the *Raven's Standard Progressive Matrices* (Raven, Raven, & Court, 1998) contains matrices that have a logical pattern or design with a missing part. The examinee selects the appropriate item that completes the pattern. Hence, progressive matrices are nonverbal; however, some researchers contend that culture also influences these skills. Another attempt to reduce cultural bias is the *Peabody Picture Vocabulary Test, Fourth Edition* (PPVT-4; Dunn & Dunn, 2007), which measures receptive vocabulary knowledge of both children and adults. There are both English and Spanish versions. In both versions, the individual is given a word and asked to simply point to the corresponding picture. The *Test of Nonverbal Intelligence, Fourth Edition* (TONI-4; Brown, Sherbenou, & Johnsen, 2010) is another example of a language-free measure of intelligence.

Conclusions

In conclusion, the information gained from individually administered intelligence tests can be helpful to counselors. The amount of useful information depends on the expertise of the individual administering and interpreting the results. Counselors need to become proficient consumers of psychological reports that include intelligence assessment. This involves having sufficient information about the instruments used to be able to evaluate the psychologist's report. For example, a counselor who receives a psychological report for a Hispanic 9-year-old boy with limited English proficiency needs to understand the cultural influences on the intelligence test used. Instrument manuals, critiques and reviews of instruments, journal articles, and books can all aid the counselor in becoming an informed professional in the area of individual intelligence assessment.

LO5 Group Intelligence Testing

Group intelligence tests are often given in conjunction with group achievement tests in schools; thus, their use is often connected with evaluating student performance. In some school settings, group intelligence tests serve as initial screening tools to identify those

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

children who should receive additional testing to identify a possible learning disability or developmental delay.

Group intelligence or ability tests can provide certain information, but they also have certain limitations. With individually administered intelligence tests, it is possible to observe behaviors that often reflect the level of motivation. In a group setting, however, it is impossible to document the behaviors of the entire group, and sometimes individuals are not motivated to perform at their highest level. Many of us have heard of children who, when taking group-administered tests, bubbled the answer sheets into attractive patterns rather than reading the questions! Furthermore, group intelligence tests require more reading than individually administered tests and are, therefore, more problematic for individuals with limited reading skills. Counselors also need to consider a client's culture, background, and language proficiencies when interpreting the results from group intelligence tests. Consistent with good assessment practices, the results of any instrument should never be interpreted in isolation, and this practice is particularly germane to group intelligence tests because they are not as sophisticated and detailed as individually administered intelligence tests.

The *Cognitive Abilities Test* (CogAT), a revision and extension of the Lorge-Thorndike Intelligence Test, is designed to assess the pattern and level of children's reasoning abilities. There are different levels of the instrument appropriate for kindergarten through grade 12. The current form of the instrument, Form 6, provides separate scores for verbal, quantitative, and nonverbal reasoning; a composite score can also be provided. The stated intent of the CogAT is to assess the level and pattern of cognitive development of students. The time allotted to take the CogAT is between 30 and 50 minutes, depending on the child's grade level. For schools interested in administering combined ability and achievement testing, the CogAT is concurrently normed with the Iowa Tests. When given with one of the Iowa Tests, the CogAT provides schools with predictive achievement scores, which can aid in identifying students who may display a discrepancy between expected achievement level and performance.

Another frequently used group intelligence test is the *Otis-Lennon School Ability Test, Eighth Edition* (OLSAT-8). The OLSAT-8 is also for students in kindergarten through grade 12 and has been re-normed. This test measures cognitive abilities related to a student's ability to learn and succeed in school. Although previous editions used the term *mental ability*, the current version does not use that nomenclature. The scores provided are Total, Verbal, and Nonverbal, in addition to five cluster scores (Verbal Comprehension, Verbal Reasoning, Pictorial Reasoning, Figural Reasoning, and Quantitative Reasoning). The maximum time requirement for the OLSAT-8 is 75 minutes. The OLSAT-8 was standardized with the *Stanford Achievement Test, Tenth Edition* (Stanford-10). Both the OLSAT-8 and the CogAT have a mean of 100 and a standard deviation of 16.

In View is a test of cognitive skills developed and concurrently standardized with the TerraNova, The Third Edition, This group test of intelligence, which can be used with students in grades 2 through 12, employs problems of sequences, analogies, verbal reasoning-words, verbal reasoning-content, and quantitative reasoning. In addition, it provides a score on the Cognitive Skills Index designed to indicate a measure of overall academic aptitude.

The *Multidimensional Aptitude Battery II* (MAB-II; Jackson, 1998) is another group intelligence test, although it is not designed for use in schools. Instead, the MAB-II was designed as a convenient alternative to the WAIS-R. The MAB-II is for either adolescents or adults and can be given either individually or in groups. Like the older Wechsler instruments, this test provides a FSIQ, a VIQ, and a PIQ. There are five verbal and five performance subtests, with each subtest having a time limit of seven minutes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

LO6 Issues in Intelligence Testing

Controversy and debate about intelligence testing are certainly not new phenomena. In fact, Gottfredson and Saklofske (2009) claimed intelligence testing has been among the most controversial topics in psychology. For example, both Jensen's (1969) oft-cited article and Hernstein and Murray's (1994) *The Bell Curve: Intelligence and Class Structure in American Life* have received enormous media attention and have ignited significant debates about the nature of intelligence and the meaning of intelligence test scores. It is important for counselors to understand some of the issues related to the construct of intelligence and the controversies surrounding the assessment of intelligence or general ability. Hence, the third section of this chapter will examine some of the issues related to intelligence testing and summarize the research findings related to these topics.

Is Intelligence Stable?

One of the major questions about intelligence test scores concerns whether they remain stable as individuals develop and age, or do scores change over time, depending on the variable influences the individual experiences in life? Infants and preschool children have the least stable intelligence test scores, a fact reflected in the lower reliability coefficients of the instruments for these younger children. The answer to whether intelligence is stable from childhood through old age depends on what research is examined.

Early research in this area indicated that intelligence tended to gradually decline after the age of 20, with the decline becoming more apparent for individuals in their 50s through their 70s. There is a commonly held notion that as we age, our memory, speed of processing, and mental agility decrease significantly. You have probably heard people cite their age as the reason for forgetting some detail. The findings of these early studies, however, have not been supported by more sophisticated research methodologies.

Early research on intellectual stability involved *cross-sectional studies*, in which people at different age levels were tested and their intelligence scores compared. What these early findings did not take into account, however, were the cultural differences that affected the scores of different generations of individuals. For example, in these cross-sectional studies, many of the older individuals did not have the same educational experiences as the younger individuals. The assessment field is in general agreement that these early results indicating decreases in intellectual function were the result of cultural differences among the different generations (Anastasi & Urbina, 1997). Later researchers were able to complete *longitudinal studies*, in which the same people were tested repeatedly at different ages in their lives. Some of the more prominent longitudinal studies (e.g., Bayley & Oden, 1955; Tuddenham, Blumenkrantz, & Wilken, 1968) indicated that intelligence gradually increases from childhood into middle age and then levels off.

Probably the best indicators of the stability of intelligence come from those studies that employed both cross-sectional and longitudinal designs, following the same individuals over time, as well as comparing 30-year-olds in 1950 with 30-year-olds in 1980 (Schaie, 1996). These well-designed studies indicated intelligence is fairly steady throughout adulthood, with slight declines occurring approximately after the age of 65. The intellectual declines that do occur tend to be in the area of fluid, as compared with crystallized, intelligence. As Verhaeghen (2003) indicated, there is little decline over the lifespan on tests that measure experience and knowledge, but there is decline on tests that require speed of processing and mental manipulation. The degree to which there may be decline in intelligence, however, appears to be related to a complex interaction of variables, such as physical health, mental activities performed in the later years, and the degree to which education is continued throughout the lifespan.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

What do Intelligence Test Scores Predict?

Earlier, this chapter presented different views concerning what constitutes intelligence and different methods for measuring intelligence. The conclusion is that there is no generally agreed upon definition of intelligence, nor is there consensus on the most appropriate methods for assessing this somewhat elusive construct. Gottfredson (2009) stressed that current intelligence tests assess relative, not absolute, levels of mental ability. Yet, intelligence tests continue to be used in our society; therefore, it is important for counselors to understand the research findings related to what these tests actually measure. The following discussion does not address findings related to any specific instrument; rather, it looks at general trends in intelligence testing.

Intelligence tests appear to be related to academic performance. The correlation between intelligence and school performance is consistently near .50 (Neisser et al., 1996). Although this is often considered a substantial coefficient, it is important to remember that only 25% of the variance is explained by the relationship between intelligence and school performance. Therefore, success in school is related to many other factors (e.g., socioeconomic factors, motivation) in addition to intelligence. However, children with higher intelligence test scores are shown to be less likely to drop out of school as compared with children with lower scores. It should, however, be noted that often school psychologists are attempting to find ways to help children learn, and a full-scale IQ score in isolation does not provide specific information that may be helpful in developing an educational plan to assist children's learning.

The relationship between IQ scores and occupational success and income is not as simple to determine. The degree to which intelligence scores are related to occupational prestige, and to a lesser degree to income, is partly due to the connection between education and intelligence. Higher-prestige occupations, such as doctor or lawyer, typically require more than a college education. Admissions scores for college (e.g., SAT) and post-degree work (e.g., GRE) are highly correlated with intelligence tests. Although Hernstein and Murray (1994) contended that intelligence scores predict occupational attainment, a task force appointed by the American Psychological Association (Neisser et al., 1996) did not reach the same conclusion. At most, the latter authors concluded that intelligence accounts for about one-fourth of the social status variance and one-sixth of the income variance. This relationship further decreased when the parents' socioeconomic status was taken into consideration.

Concerning the relationship between IQ scores and occupational success, the research conducted by Hunter and Schmidt (e.g., Hunter, 1986; Hunter & Schmidt, 1983) is often cited because they studied the effects of abilities and aptitudes on job performance. The findings, which Hunter and Schmidt termed **validity generalization**, are primarily related to aptitude testing and the *General Aptitude Test Battery* (GATB), a vocational aptitude test developed by the U.S. Department of Labor. The GATB does include a measure that some professionals contend is synonymous with intelligence; thus, Hunter and Schmidt's findings do have some relevance in regard to the predictive ability of intelligence tests.

Hunter and Schmidt, and their associates, were primarily interested in studying which variables predicted successful personnel selection. In studying the effectiveness of tests in employment selection, they focused on the GATB because numerous validation studies had been conducted. Other researchers contended that the GATB was fine at predicting performance in some occupations but not good at predicting performance in other occupations. By examining the variations in validity coefficients of job performance tests, which primarily involved the GATB and performance in various occupations, they found that differences in validity coefficients were mainly due to measurement and sampling error (Pearlman, Schmidt, & Hunter, 1980; Schmidt & Hunter, 1977; Schmidt, Hunter, & Caplan, 1981). In a large meta-analysis, Hunter (1980) found that the three composite scores of

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the GATB were valid predictors of success in all jobs. He also found that the relationship between the aptitudes and job performance was linear, indicating that as test scores increased so did ratings of job performance. Hence, the best workers had the highest test scores. Thus, the term *validity generalization* refers to the findings that the same test score data may be predictive for all jobs. The implication of this research is that if a test is valid for a few occupations, the test is valid for all jobs in that cluster. Therefore, these researchers found evidence that validity coefficients could be generalized to other occupations.

More recent research tends to support the hypothesis that cognitive ability testing predicts work performance (Kuncel, Ones, & Sackett, 2010). In fact, Ones, Viswesvaran, and Dilchert (2004) found that the validity of cognitive ability tests in predicting overall job performance is about .50. This finding comes from over 20,000 studies and included data from over 5 million participants. Similar to academic performance, this means that only 25% of the variance is shared variance, so many other factors are also related to work performance (e.g., motivation, personality).

Is Intelligence Hereditary?

This is one of the most controversial issues in intelligence testing, for the question is essentially asking how much of measured intellectual ability is innate and how much is due to environmental factors. There are many traits that vary among people (e.g., weight, propensity for some diseases, or aspects of personality), and scientists have attempted to discern the amount of variation that is attributable to genetics and the amount that is due to environmental factors. Gottfredson and Saklofske (2009) contended that it is well established that intelligence is a combination of both heredity and environment and the interaction of the two. A *heritability index* (h^2) is intended to provide an estimate of the proportion of variance associated with genetic differences. The reverse then applies, with $1 - h^2$ being an indicator of the amount of variance due to environmental factors. Determining estimates of the genetic contribution to intelligence is, at this point, difficult. Some of the studies have examined the differences between monozygotic (MZ) twins and dizygotic (DZ) twins who have been raised together and those who were adopted and raised apart. MZ twins have the same genetic makeup. Other studies involved several kinds of kinship and the influences of genetics on intelligence.

In general, the heritability indexes for intelligence tend to be approximately .50 (Chipuser, Rovine, & Plomin, 1990; Loehlin, 1989). McGue, Bouchard, Iacono, and Lykken (1993) argued that this h^2 figure would probably be larger (i.e., approximately .80) if researchers only included studies with adult samples. Although genes appear to have a substantial influence on performance on intelligence tests, this effect is not sufficiently large enough to indicate that intelligence is inexorably set at conception. Environmental factors also have a significant effect on intellectual development. Neisser et al. (1996), in their extensive review of research in this area, concluded that variation in the unique environments of individuals and between-family factors significantly influence IQ scores of children. IQ scores seemed to be most related to the interaction between an individual's genetic makeup and environmental influences. A child is born with certain capabilities, but the degree to which those capabilities are realized is significantly influenced by the environment surrounding that child.

What Environmental Factors Influence Intelligence?

Many environmental factors influence the intellectual development of an individual, but it is clear that the *cultural* environment in which a person matures has a significant influence on that person's intellectual development. Postmodern philosophy has stressed the influences of culture and language in the overall development of an individual (Gergen, 1985). A child

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

of Indian descent growing up in rural Malaysia has different experiences than an upper middle–class Caucasian child growing up in San Francisco. Cultures differ in a multitude of ways that are both overt and subtle; therefore, we are unable to directly identify all the cultural factors that influence performance on intelligence tests. Even within a population, there are subgroups with cultural variations, such as the many different Native American tribes in the United States. For example, the culture of the Arapahos is different from the culture of the Shoshones, even though the two tribes share a reservation in Wyoming.

Although intelligence tests try to eliminate the effects of schooling, attendance at school does influence performance on intelligence tests (Ceci, 1991). Children who attend school regularly have higher test scores than children who attend intermittently. If we examine differences between children who are of the same age but, because of birthday deadlines, vary in the number of years attending school, those who have attended longer will have higher IQ scores. Precisely how schooling influences intelligence test performance is difficult to determine. Certainly, many intelligence tests incorporate information items such as "In what country were the first Olympic games?" The probability of being exposed to this information increases with attendance at school. Children in school are also more likely to be exposed to manipulation of objects and puzzles. Furthermore, schools promote problem-solving skills and abstract thinking. It must be noted that not all schooling is alike and that the quality of the school experience also influences IQ scores. Those who attend poor-quality schools will not receive the same instruction, and their intelligence scores will reflect this deficiency (Neisser et al., 1996).

Family environments have an influence on many aspects of development, including cognitive development. Early research indicated that intelligence scores were directly influenced by parents' interest in achievement (Honzik, 1967) and methods of disciplining (Baldwin, Kalhorn, & Breese, 1945; Kent & Davis, 1957). Research has since questioned these early findings (Scarr, 1992, 1993). Severely neglectful or abusive environments, however, can negatively affect children's cognitive development. For example, prolonged malnutrition in childhood impedes cognitive development. After a child is provided a minimally nurturing environment, it is difficult to determine what other family factors may contribute to performance on intelligence tests.

Environmental factors that can affect intelligence also include exposure to certain toxins. The negative consequences of exposure to lead have been well documented (Baghurst et al., 1992; McMichael et al., 1988). Prenatal exposure to large amounts of alcohol can produce deleterious effects on the child, termed *fetal alcohol syndrome* (Stratton, Howey, & Battaglia, 1996). Cognitive delays are associated with fetal alcohol syndrome. There is conflicting research about the effects on cognitive development related to prenatal exposure to small amounts of alcohol, caffeine, over-the-counter pain medication, and antibiotics.

Are There Group Differences in Intelligence?

Probably the most controversial issue related to the testing of intelligence concerns ethnic and gender group differences. Before we begin a discussion of these group differences in intelligence, it is important to remember that research has continually shown that within-group variance is greater than between-group variance. This means that within any ethnic group, the variation in scores is quite large; in fact, the variation *within* the group is greater than *between* groups. Hence, there is overlap among the groups that challenges the conclusions about the superiority or inferiority of one group compared with another. Furthermore, because of the large variation in intelligence scores, it is inappropriate to stereotype any individual based on group performance scores.

Are There Gender Differences in Intelligence? Although there do not appear to be general intellectual differences between males and females (Spinath, Eckert, & Steinmayr, 2014), there are some differences on specific tasks or abilities. Men appear to be better

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

in visual-spatial ability, which involves tasks that require the individual to visualize and mentally rotate objects. In a meta-analytic study, Masters and Sanders (1993) found an effect size of .90, indicating that men in general score almost a standard deviation higher than females on visual-spatial tests. On the other hand, females appear to have the advantage over males in terms of some verbal tasks. On synonym generation and verbal fluency, females score higher than males, with the effect sizes ranging from .50 to 1.20 (Gordon & Lee, 1986; Hines, 1990). Thus, females' scores on tests of verbal tasks tend to be between a half a standard deviation to more than a full standard deviation higher than those of males.

Are There Ethnic Group Differences in Intelligence? In general, African Americans and Hispanics tend to score lower on intelligence tests than European Americans or Asian Americans. African Americans and Hispanics tend to score about 10 points below Caucasians (Weiss, Saklofske, Prititera, & Holdnack, 2006). Socioeconomic factors seem to have some influence on test scores; however, black/white differences are reduced, although not eliminated, when socioeconomic factors are controlled for. Chapter 6 addresses questions of test bias, which some researchers have argued contributes to the disparity in IQ means between African Americans and Caucasians. Some research does indicate that the difference between African Americans and Caucasians on IQ scores is decreasing, but more definitive studies are needed before this can be considered an established finding (Neisser et al., 1996).

Hispanic Americans' average intelligence test scores are typically between those of African Americans and Caucasians (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). Counselors need to be aware that linguistic factors may influence Hispanic individuals' performance on intelligence tests. Hispanic children often score higher on the performance subtests than on the verbal subtests. For many Hispanic children, English is not their first language, which may partially explain the discrepancy between performance and verbal scores. Dana (2005) contended that translation of standard intelligence tests into Spanish does not resolve these problems.

Concerning Native Americans' performance on intelligence tests, counselors need to be careful not to generalize due to the multitude of distinct tribes within this population. The traditions and cultures of the Native American tribes vary; thus, performance on aspects of intelligence tests also tends to vary. A general finding is that Native American children, like Hispanic children, seem to score lower on the verbal scales as compared with the performance scales (Neisser et al., 1996). In addition to cultural differences, McShane and Plas (1984) suggested that Native American children are also plagued by chronic middle-ear infections that can negatively affect their development in the verbal area. The behavior of Native Americans during testing may also be misinterpreted by examiners. Native Americans tend to be deferential to a Caucasian examiner, which examiners may perceive as a lack of motivation.

There is some debate concerning the average intelligence scores for Asian American children, with studies reporting averages from around 100 to 110. A general finding is that Asian Americans achieve substantially more, both academically and occupationally, than their intelligence scores would predict. Asian Americans tend to do well in school, and a higher proportion of them are employed in managerial, professional, and technical occupations than would be expected based on their intelligence scores. This "overachievement" demonstrates some of the problems with IQ-based predictions (Neisser et al., 1996). It should be noted that there is also great diversity among Asian Americans (e.g., Chinese, Korean, and Vietnamese).

The ethnic group difference in intelligence test performance does not appear to be related to any one factor, such as heredity or language differences. Neisser et al. (1996) concluded that researchers have not yet been able to explain these differences. Sternberg and Grigorenko

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(2008) argued that the research related to group differences is rooted in the model one uses to interpret the relationships between general ability, intelligence tests, and culture. Sternberg and Grigorenko contended there are four models from which scholars and practitioners examine ability testing across cultures. In the first model, the theoretical position is that the nature of intelligence is precisely the same cross-culturally and that when appropriate accommodation are made (e.g., translation) then the same intelligence tests can be used across different cultures. The second model is that there are differences in the nature of intelligence cross-culturally but the same measures are used to measure it. With this model, the focus is on the different outputs and understanding the structural differences in intelligence based on the different cultures. The third model, which is the one Sternberg and Grigorenko support, is that the dimensions of intelligence are the same across cultures but the measures should not be the same. Sternberg and Grigorenko contended that general ability concerns developing expertise and the specifics of developing expertise vary across cultures and, therefore, the intelligence tests for assessing the abilities to gain expertise need to correspond to the culture. The fourth model is that the nature of intelligence is different across cultures and, therefore, intelligence tests need to correspond to the nature of intelligence.

In conclusion, ethnic group difference in intelligence test performance does not appear to be related to any one factor, such as heredity or language differences, and it may be related to the theoretical lens from which one views the research. In counseling individuals, counselors need to remember that there is great variation in intelligence within all ethnic groups and that a counselor cannot predict a client's level of cognitive abilities based on ethnicity. Stereotypes concerning any individual based on his or her ethnic, gender, or any type of group can be detrimental.

What Is the Flynn Effect?

As this section of the chapter is devoted to issues in intelligence testing, it is important to note the steady rise of intelligence scores in recent years. This rise in intelligence score is often called the "*Flynn effect*" after James Flynn (1984, 1987), who was one of the first researchers to identify this trend. In the past 50 years, the average IQ score has increased more than a full standard deviation. The average gain is about three IQ points each decade. There are numerous possible explanations for the Flynn effect, such as more test sophistication, changes in education and educational opportunities, and changes in parenting practices. Although the increases in intelligence scores are well documented, the reasons for this rise are debated within the field. There are also differing opinions about whether this trend will continue or if there will be a leveling-off of average IQ scores. It should also be noted that in some countries, the average IQ scores are not continuing to increase (see Flynn, 2013).

Summary

Intelligence is a word used in everyday conversation (e.g., "John is intelligent" or "Someone is not very intelligent"). In addition, intelligence testing has a rich and substantial history, with certain instruments having been used extensively in a variety of settings. Given these factors, one might expect more agreement on the definition and theoretical basis for intelligence. Numerous theorists contend that a psychometric or factorial approach best explains the construct of intelligence. These individuals suggest that intelligence is composed of one or more factors. There is some debate about whether intelligence can best be described as one general factor (*g*) that affects all cognitive performance. Many intelligence tests have a composite score or a full-scale IQ score that represents a measure of *g*.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Other factor theorists contend that intelligence can be better explained by focusing on lower-order factors, such as the differences between fluid and crystallized intelligence. Currently, the Cattell-Horn-Carroll's is having the most influence on the testing of intelligence or general ability. Developmental theorists suggest that intelligence can be better understood by examining how the structure of intelligence develops. Certainly, the most influential theorist in this developmental area is Piaget. Rather than studying the structures or development of intelligence, Luria's information-processing model concerns how individuals perceive information and make sense of that information. A number of researchers currently hope that by analyzing how the brain works, they will understand the phenomenon of intelligence. Finally, there are new models of intelligence that propose expanding the methods of testing intelligence. The models of Sternberg and Gardner, for example, are influencing how the field considers and conceptualizes intelligence.

Theories of intelligence have influenced the development of some of the most commonly used intelligence tests. To assess a person's cognitive capabilities, an individually administered intelligence assessment may be conducted. The most frequently used instruments are those developed by Wechsler. The various Wechsler intelligence tests are similar in content, but each has been specifically designed for and standardized for a specific age range. The *Wechsler Adult Intelligence Scale—IV* (WAIS-IV) is for adults, the *Wechsler Intelligence Scale for Children—V* (WISC-V) is for children ages 6 years, 0 months through 16 years, 11 months old, and the *Wechsler Preschool and Primary Scale of Intelligence—IV* (WPPSI-IV) is for children ages 2 years, 6 months through 7 years, 7 months old. Another historically significant scale, the Stanford-Binet, is currently in its fifth edition. The authors of the Kaufman Assessment Battery for Children, Second Edition (KABC-II) designed their test to be appropriate for children from diverse ethnic groups. Another battery of instruments that is often used is the Woodcock-Johnson IV Complete Battery, which can include an individual assessment of general ability, achievement, and oral language. Although group intelligence tests are easier to administer, they cannot provide the degree of information that individually administered intelligence tests provide. Furthermore, performance on group intelligence tests is highly dependent on the examinees' reading abilities. It also should be noted that cognitive or general ability tests are less reliable with younger children than they are with adolescents and adults.

Numerous issues relate to the use of intelligence tests. Information from intelligence tests can be helpful in counseling clients. A skilled clinician, however, must be aware of the strengths and limitations of these measures. Multicultural factors are particularly important to consider when results from intelligence tests are used in decision making. Intelligence test results can provide information about a client's cognitive abilities, but that information needs to be integrated with other contextual factors to give a more thorough understanding of the client.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 10

Assessing Achievement and Aptitude: Applications for Counseling

Learning Objectives

- LO1 Describe the difference between achievement and aptitude testing
- LO2 Understand the influence of the No Child Left Behind legislation on achievement testing
- LO3 Identify the Common Core State Standards and recognize their influence on achievement testing
- LO4 Understand the use of the National Assessment of Educational Progress (NAEP)
- LO5 Realize the uses of individual achievement tests
- LO6 Describe the major scholastic aptitude tests
- LO7 Explain the major vocational aptitude tests
- LO8 Distinguish between test sophistication skills and coaching

Jenna is a 17-year-old high school senior who is unsure about her college plans. Her parents both graduated from Ivy League schools and want Jenna to attend one of those institutions. Jenna had planned on going to an Ivy League school but is now wondering if she is smart enough. Jenna indicated that she was involved in many high school activities in her freshman, sophomore, and junior years; however, she has dropped out of these activities during her senior year. In addition, Jenna and her mother have been arguing about Jenna's "lack of motivation." Jenna's grades continue to be mostly As with some Bs. Her parents have insisted she go to counseling to improve her attitude and get back to preparing for a "good" college. Jenna has provided you, her counselor, with her scores on both the SAT and the ACT. Her SAT Reasoning Test scores were 570 for the Evidence-Based Reading and Writing section, 590 for the Math score, and 500 for the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Writing section. On the ACT, Jenna received 22 in English, 24 in Mathematics, 21 in Reading, 23 in Science, and a Composite score of 23. Jenna and her parents want your advice on whether she is "bright enough" to go to a selective college. Based on these scores, what might you tell them? In addition, what other information might you collect that would be helpful in providing feedback to Jenna about her academic potential? Jenna also states that she has a severe case of test anxiety and wants help in improving her scores. Hence, you are considering interventions that might help her test anxiety and wondering if these interventions could significantly improve her performance.

LO1 **As you will remember** from Chapter 1, achievement tests provide information about what individuals have learned or knowledge they have acquired; whereas, aptitude tests predict future performance. Jenna's case is but one example of an instance when it would be helpful for a counselor to have an understanding of achievement and aptitude assessment. Another example is the situation of a second-grade boy whose teacher is concerned because he does not appear to be reading at grade level. The teacher is unsure of what may be contributing to the difficulties in reading as his parents have recently separated and have indicated to the teacher they will soon be getting divorced. The results of an achievement test may help both the teacher and the school counselor know if the boy is having difficulties with reading and, depending on the test, may even indicate where some of the difficulties are related to the acquisition of reading skills.

Knowledge of aptitude assessment can also be useful to counselors. Aptitude tests predict future performance or ability to learn. Clients often come to counseling because they are trying to make a decision about their futures (e.g., seek a new job, go back to school, decide on a career), and aptitude tests can provide some useful information in these decisions. In fact, many of us have probably wondered at some point in our lives if we had the aptitude for a particular task. Many aptitude tests are used for selection purposes, such as employers selecting employees and college admissions personnel deciding which students should be admitted to certain colleges. Some of these same aptitude assessments, however, can be used in counseling situations to assist clients in making better-informed decisions about their futures. In general, achievement tests measure acquired learning, whereas aptitude tests measure potential or ability to learn new tasks. Although there are a number of different types of both achievement and aptitude tests, sometimes the differences between them are somewhat ambiguous. The contrast between achievement and aptitude tests is one of purpose more than of content. For example, the same math problems might be used to determine if someone has learned a mathematical concept (achievement) or to predict future performance in jobs requiring some mathematical knowledge (aptitude). This chapter briefly discusses some of the achievement and aptitude tests commonly used by counselors.

Assessment of Achievement

In order to understand current issues in achievement testing, it is important to examine what has recently been happening in public education and to understand some of the legislation that has influenced achievement testing. With increased scrutiny on student learning, there has been increased interest in achievement testing. Achievement testing is not only used to assess individual students, but they are also being used related to educational accountability and whether schools and teachers are performing adequately. Furthermore, achievement testing is currently a dynamic process that is continually changing and evolving. New state or federal laws, regulations, and guidelines are dictating changes in achievement tests, and how achievement test results are being used is changing.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

One of the major issues in achievement testing is the increases of standardized achievement testing in all 50 states (Tindal, 2002). This testing has sometimes been labeled **high-stakes testing** because of how the assessment information may be used. Although some states began instituting "high stakes" in the early 1990s, with the passage of the No Child Left Behind Act (NCLB) of 2001, all states needed to institute an assessment program. The Obama administration released its blueprint for revising the Elementary and Secondary Education Act (ESEA) on March 13, 2010 (NCLB was a reauthorization of the Elementary and Secondary Education Act of 1965). As of the writing of this book, however, the reauthorization of the ESEA has not occurred, but it is anticipated that the new law will continue to focus on educational accountability.

LO2 No Child Left Behind

In order to understand current practices in achievement practices, it is important for the reader to have some understanding of the No Child Left Behind Act or the reauthorization of the Elementary and Secondary Education Act of 1965. The purpose of NCLB was clearly stated in its introduction, "The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments" (SEC, 1001). Central to NCLB was a focus on educational accountability by examining the progress of individual students (Dollarhide & Lemberger, 2006). In terms of accountability, the law stipulates that all students in grades 3 through 8 (not just Title I students) must be assessed annually in at least math and reading so that their progress can be measured against state standards. Each state had to develop measurable adequate yearly progress (AYP) objectives to ensure that students had reached the determined objectives. Furthermore, there had to be objectives for specific groups, such as economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and students with limited English proficiency. The adequate yearly progress objectives were measured at the school level, and schools that did not meet the state objectives for two years were identified as needing improvement. When a school failed to make adequately yearly progress for two years, parents of children attending that school were permitted to transfer to another school in the district. After not making adequate yearly progress for three years, the school had to provide supplemental educational services. Failure to make adequate yearly progress for years 4 and 5 led to additional sanctions that needed to constitute a major restructuring of the school, such as instituting a new curriculum and replacing staff. Eventually, failure to make adequate yearly progress led to a takeover of the school by the state or the state could contract with a separate agency to run the school.

The states were given 12 years (i.e., the end of the 2013–2014 school year) to show that 100% of their students are performing at the proficient level or above in mathematics and reading/English. As this deadline as past, this further drives the reauthorization process. Furthermore, it must be acknowledged that a 100% proficiency is a high bar. In September 2011, the U.S. Department of Education invited each State educational agency to request flexibility regarding specific requirements of the No Child Left Behind Act of 2001 in exchange for rigorous and comprehensive State-developed plans designed to improve educational outcomes for all students. Thus, educational accountability systems are developed at the state level, but these systems must be approved by the U.S. Department of Education. The U.S. Department of Education requires that these accountability systems be rigorous and comprehensive with a focus on closing the achievement gaps, increasing equity, and improving the quality of instruction.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In providing guidance to the State education agencies (SEA) regarding flexibility while remaining incompliance with NCLB, the U.S. Department of Education requires that each state have expectations that all students graduate from high school ready for college and a career, by implementing State-developed college- and career-ready standards. In addition, each SEA has to implement high-quality assessments aligned to those standards. Furthermore, SEA are required to implement accountability systems for schools and districts that incorporate a variety of measures, including rigorous performance and graduation rate targets, to identify underperforming schools. An SEA also has to be committed to implement teacher and principal evaluation and support systems based on multiple measures, which need to include student growth as a significant factor. This typically involves using standardized achievement tests to measure student growth.

Although each state is determining how they are providing systematic and rigorous assessment of students in grades 3 through 8, there is still concern about how these achievement test scores are used. Some schools and states have used their state's assessment to meet NCLB to make placement decisions (e.g., remedial class, grade promotion). When a test is used to make a major decision (e.g., if a child can graduate from high school), this is often labeled high-stakes testing. The American Psychological Association (2001) has taken a stand on this type of testing:

> *Measuring what and how well students learn is an important building block in the process of strengthening and improving our nation's schools. Tests, along with student grades and teacher evaluations, can provide critical measures of students' skills, knowledge, and abilities. Therefore, tests should be part of a system in which broad and equitable access to educational opportunity and advancement is provided to all students. Tests, when used properly, are among the most sound and objective ways to measure student performance. But, when test results are used inappropriately or as a single measure of performance, they can have unintended adverse consequences. (p. 1)*

The American Educational Research Association (2000) also issued a position statement concerning high-stakes testing in pre-K–12 education. This statement is based on the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999). Central to AERA's (2000) argument is the premise that decisions that affect an individual student's life chances or educational opportunities should not be based on one test score, which is consistent with APA's position that was mentioned earlier. Furthermore, AERA contended that all high-stakes achievement testing programs in education should meet the conditions of (1) adequate resources and opportunities to learn; (2) validation for each separate intended use; (3) full disclosure of likely negative consequences of high-stakes testing programs; (4) alignment between the test and the curriculum; (5) validity of passing scores and achievement levels; (6) opportunities for meaningful remediation for examinees who fail high-stakes tests; (7) appropriate attention to language differences among examinees; (8) appropriate attention to students with disabilities; (9) careful adherence to explicit rules for determining which students are to be tested; (10) sufficient reliability for each intended use; and (11) ongoing evaluation of intended and unintended effects of high-stakes testing. In conclusion, it is clear that achievement testing will continue to have a major role in educational decisions and policies and that the debates about high-stakes testing will continue.

In concluding our discussion of NCLB, it is important to remember that the reauthorization of the ESEA will make changes in current accountability systems. It is difficult to predict the precise changes as there are varying opinions of the effectiveness and fairness of NCLB. In my opinion, there will be a continued focus on educational accountability and an emphasis on measuring student learning. Hence, assessment, particularly achievement testing, will continue to play a prominent role in education in the United States.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L03 Common Core

In order for counselors to understand current trends in achievement testing, it is also important that they understand the rise of educational standards in public education and the **Common Core State Standards**. Although some people believe the Common Core was developed by the federal government, this is not accurate. The Common Core was launched in 2009 by state leaders, including governors and state commissioners of education from 48 states, two territories, and the District of Columbia, through their membership in the National Governors Association Center for Best Practices and the Council of Chief State School Officers. The rationale for developing the Common Core was to develop rigorous, consistent, and real-world learning goals that would ensure that all students, regardless of where they lived, would graduate from high school prepared for college and career.

The development of the Common Core State Standards was influenced by many sources of information including existing quality state standards and experiences of teachers, content experts, state leaders, and leading researchers. The development of the Common Core standards involved many working groups, various committees, and extensive public comments. For example, the National Governors Association Center for Best Practices and the Council Chief State School Officers (CCSSO) received nearly 10,000 comments on the standards during two public comment periods. The development of the Common Core standards was a two-step process where *college- and career-readiness standards* were first developed. These standards clearly state what students were expected to know and understand by the time they graduate from high school. Then working backward, the various committees and working groups developed the *K-12 standards*, which systematically articulate what students need to know at each grade level in order to meet the college- and career-readiness standards. It should be noted that the Common Core stipulates standards only in the areas of English/language arts and mathematics.

The Common Core is produced by the National Governors Association Center for Best Practices and the Council Chief State School Officers, and it has been approved in 43 states, the District of Columbia and four territories and by the Department of Defense. The Common Core does not stipulate how students are taught nor does it specify what measures are used to assess student learning. However, one of the rationales for developing the Common Core is so that states could pool their resources and develop strong and valid measures to assess students on a common set of learning outcomes.

Assessment of the Common Core

As the Common Core has been approved in 43 states, the District of Columbia, and four territories and by the Department of Defense, it seems important for counselors to be familiar with the typically used assessments in these states and territories. Early in the process of developing the Common Core, two consortia emerged to develop assessments to assess the common core. These two consortia are the *Smarter Balanced Assessment Consortium* and *Partnership for Assessment of Readiness for College and Careers* (PARCC). I have selected Smarter Balanced as an example of achievement testing that is designed to correspond to the Common Core.

Smarter Balanced is a comparatively new assessment as the field testing of the instrument was conducted in spring 2014. All 21 governing states and the U.S. Virgin Islands participated in the field testing. Smarter Balanced will be implemented in states for the first time during the spring of 2015. Smarter Balanced includes both summative and interim assessments and formative resources for teachers. The interim assessment is voluntary, and it is designed to forecast how students will do on the summative assessment. There are also online practice tests for students to take. The summative assessments are designed for grades 3 through 8 and grade 11.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The summative assessment is given during the last 12 weeks of the school year. The summative assessments for each grade level consist of two parts: a computer-adaptive section and a performance section. With **computer-adaptive testing (CAT)**, the computer adapts the next question for the student based on his or her response. For example, if a student answers the item correctly, then he or she will receive a more challenging item next, whereas, if the student answers the item incorrectly, then he or she will receive an easier item. The performance section of Smarter Balanced is not computer adaptive, and it requires the students to apply their knowledge and skills to respond to complex real-world problems. The performance section is also taken at the computer.

Smarter Balanced has developed content specifications in English language arts/literacy and mathematics to ensure that the assessments cover the range of knowledge and skills in the Common Core. Various work groups contributed to these content specifications, and they were also open for public comment. Item were written with sensitivity to bias, accessibility, and accommodations. Smarter Balanced uses an evidence-centered assessment design, which is an approach in education assessment where the focus is evidentiary arguments built upon intended constructs, with explicit attention paid to the potential influence of unintended constructs. For readers interested in evidence-centered assessment design, see Mislevy, Steinberg, and Almond (2003).

As state and even some schools within states have the choice of assessments, there are other assessments of the Common Core that have developed outside of the two consortia. For example, ACT has developed ACT Aspire, which can be taken either online or in a paper format. Like Smarter Balanced, ACT Aspire has both summative and interim assessment options. ACT Aspire is anchored with the commonly known ACT assessment, which is designed to predict college performance and will be discussed later in this chapter. ACT Aspire is intended for students in grades 3 through 8 and early high school (grades 9 and 10) and assesses in the subject areas of English, math, reading, science, and writing.

For students who take ACT Aspire in proceeding years, their scores are documented on the report, which includes their student growth percentiles. ACT Aspire plots scores for each of the subtests (i.e., English, math, reading, science, and writing) and a composite score. These scores are compared to the ACT Readiness Benchmarks, which indicates whether the student is on track for college and career readiness. Furthermore, ACT Aspire also provides predicted score ranges for the next two years for each of the five subject areas as well as the composite score range. Therefore, teachers, parents, and counselors can see students' expected educational growth. For grade 9, ACT Aspire reports include predicted grade 10 ACT Aspire score ranges and predicted grade 11 ACT score ranges; whereas, for students who take ACT Aspire in grade 10, the prediction is just to grade 11 ACT scores.

Smarter Balanced and ACT Aspire are just two examples of assessments designed to correspond to the Common Core State Standards, and there are other assessments that are used in different states. Also, not all states have adopted the Common Core, but states are still required to assess students in grades 3 through 8 in at least the areas of English/language arts and mathematics. For states not a part of the Common Core, they will need to have assessments that correspond to their state's educational standards.

LO4 National Assessment of Educational Progress

As states may vary in their performance standards and state tests could vary in their rigor, a sample of students in grades 4 and 8 will be required to take the National Assessment of Educational Progress (NAEP) in reading and mathematics every few years. The NAEP, also known as "the Nation's Report Card," is the only national assessment of students' knowledge and performance in various subject areas. Since its inception in 1969, it has been used periodically to assess student achievement nationally in reading, mathematics,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

science, writing, U.S. history, civics, geography, and the arts. The NAEP does not provide evaluation of individual students or even schools because the original intent of the NAEP was to report on the achievement of the nation, to examine performance of demographic subgroups, and to track trends in educational progress over time. Only since 1990 has the NAEP been used to provide feedback to participating states. Before the NCLB legislation, the purpose of NAEP was to survey what students knew and were able to do. State and school participation was voluntary. After NCLB, the NAEP assessments have taken on a new role, and states and school districts, if selected, must participate in reading and math assessment every two years. It should be noted that when states are assessed, only a sample of students will be required to participate. The student sample for a state is drawn to be representative of that state, and minority students and students attending private schools are oversampled to provide a sufficient sample size for mandated subgroup analyses. The NAEP uses both objective and constructed-response items. Sample questions are available to school districts. After each administration of the assessment, about 30% of the items are replaced. Although item pools are not identical from year to year, item response theory is used to ensure comparable items and to ensure that subject area and content subdomains are assessed consistently over time.

Adult Achievement Testing

There are also achievement batteries available for adults because some adults seek information about their current level of achievement in order to upgrade their skills. For example, a high school dropout may want to know where his academic deficiencies are so that he can learn those skills that were missed. An example of an adult achievement battery is the *Test* of *Adult Basic Education* (TABE), which includes both a Survey Version and a Complete Battery. The Survey Version can be used for screening and placing, whereas the Complete Battery provides much more thorough information for determining educational interventions. The TABE Complete Battery, which is designed to correspond to the *General Education Development* (GED) test, assesses the academic areas of vocabulary, reading, language, language mechanics, mathematics, and spelling. The TABE can also be used for other purposes and indicates whether an adult has made progress in literacy education or learning English. The TABE is also used in programs related to workforce development because it measures work-related basic skills and work-related problem-solving skills. The different levels of the TABE (i.e., Limited Literacy, Easy, Medium, Difficult, and Advanced) enable examiners to assess the basic skills of adults at various levels of literacy.

LO5 Individual Achievement Tests and Diagnostic Achievement Tests

Another type of achievement tests concerns those instruments that are administered individually. Some of these achievement instruments are appropriate for adults and can be used to provide retraining for adults (e.g., vocational rehabilitation) to assess the academic skills a person has retained. The most prevalent use of individual achievement tests, however, is for psychoeducational evaluations in which children are screened for learning disabilities, cognitive handicaps, or behavioral disorders, or for other academic referrals (Sattler, 2008).

In the past, a common method for diagnosing a learning disability was to look at the discrepancy between intelligence and achievement. This method of diagnosing a learning disability typically involved the use of an individually administered intelligence test (e.g., WISC-V) and an individually administered achievement test. With the discrepancy model, a learning disability was indicated when a person was not achieving at a level consistent with his or her abilities or intelligence. Previously, if there was a difference of two

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

standard deviations between the achievement level and intellectual level, then a learning disability was diagnosed. The 2004 reauthorization of the Individuals with Disabilities Education Act (IDEA), however, explicitly states that schools are *not* required to use a severe discrepancy between achievement and intellectual abilities to find that a child has a specific learning disability. The replacement for the discrepancy approach of diagnosing a learning disability is **response-to-intervention** (RTI).

The focus of RTI is for schools to use data (e.g., achievement tests, classroom activities) to identify students at risk for poor learning outcomes. Teachers then monitor student progress and provide evidence-based interventions. Progress is then monitored by school personnel, and a multileveled prevention system is available for students who are not making adequate progress. Adjustments to the intensity and nature of those interventions are made, and student progress is continued to be monitored. Within the RTI framework, there are three levels of prevention: primary, secondary, and tertiary. The primary level is within the classroom; the secondary level is evidence-based intervention(s) of moderate intensity that addresses the learning or behavioral challenges of most at-risk students. Therefore, with a child who appears to be behind in an academic area or who is not performing at an expected level, primary prevention is tried first within the classroom. If modifying classroom activities does not solve the problem, then secondary prevention activities are implemented. These secondary activities must be evidence-based, and the child's progress must be monitored. If sufficient progress is still not achieved through evidence-based instruction, then the student's case is moved to the tertiary level. At the tertiary level, there are individualized interventions for students who show minimal response to secondary-level interventions. This is the level in which an individual achievement test may be used both to monitor progress and to possibly diagnose a student with a learning disability. RTI integrates assessment with instructional interventions (National Center on Response to Intervention, 2010).

The Kaufman Test of Educational Achievement, Third Edition (KTEA-3 Comprehensive Form; Kaufman & Kaufman, 2014) is an individually administered assessment of achievement for grades pre-kindergarten through 12 or ages 4 through 25 years. There are two sets of norms: one based on grades, which includes children from pre-kindergarten through grade 12, and second, one based on ages, which includes individuals from age 4 through 25 years, 11 months. Individuals can be assessed with the KTEA-3 Comprehensive Form in close proximity as there are two independent, parallel forms (A and B). The KTEA-3 can be used as a comprehensive measure of achievement where the intent is to identify academic strengths and limitations or it can be used to narrow in on a specific area of achievement (e.g., reading). The KTEA-3 measures achievement in reading, mathematics, written language, and oral language.

The KTEA-3 is frequently used to diagnose learning disabilities, and it provides measures of all eight of the specific learning disability areas covered by IDEA 2004 and the three learning disability diagnoses in the *DSM-5*. It is also mapped to the Common Core State Standards. There are various subtests and composite scores that can be used to identify learning disabilities. Furthermore, the error analysis system provides insight into instructional interventions by allowing the examiner to compare an examinee's pattern of errors to a normative reference group. These instructional interventions are included in the computerized report that can be purchased through the publisher. The KTEA-3 content has undergone extensive bias reviews and statistical analyses to ensure fairness. It can also be used in concert with the Kaufman Assessment Battery for Children, Second Edition (KABC-II) and the Wechsler Intelligence Scale of Children, Fifth Edition (WISC-V).

There is also a major individual achievement test that is associated with a family of well-respected instruments—the *Wechsler Individual Achievement Test-Third Edition* (WIAT-III; Wechsler, 2009). The WIAT-III can be used with individuals from 4 years old

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

through adulthood. The WIAT-III contains 16 subtests, although not all subtests contribute to the Total Achievement composite. The 16 subtests are Listening Comprehension, Early Reading Skills, Reading Comprehension, Math Problem Solving, Alphabet Writing Fluency, Sentence Composition, Word Reading, Essay Composition, Pseudoword Decoding, Numerical Operations, Oral Expression, Oral Reading Fluency, Spelling, Math Fluency-Addition, Math Fluency-Subtraction, and Math Fluency-Multiplication. Several derived scores are available for each subtest, and these can be used in the diagnosis of specific learning disabilities. The composite scores may also be used. These composite scores are Oral Language, Total Reading, Basic Reading, Reading Comprehension and Fluency, Written Expression, Mathematics, Math Fluency, and Total Achievement. For both the subtest and the composite scores, growth scale values are available so that progress can be monitored. Miller (2010) contended that the WIAT-III can be used to inform decisions about eligibility for services, placement, and diagnosis of specific learning disabilities; however, he asserted that it does not provide evidence (reliability or validity) for the suggested use of the test to design instructional objectives and plan interventions.

Another individual achievement test that is based on earlier versions is the *Wide Range Achievement Test 4* (WRAT4; Wilkinson & Robertson, 2007). The WRAT4 assesses achievement in the areas of reading, spelling, and arithmetic. The WRAT4 was developed to assess the basic academic skills of individuals ages 5 through 94. Consistent with other individually administered achievement tests, Robertson (2002) developed an expanded version of the WRAT, and it is anticipated that an expanded version of the WRAT4 will also be developed. The current WRAT-Expanded has two forms: Form G for group administered assessment and Form I for individually administered assessment.

In the assessment of possible learning disabilities, psychologists might also use content-specific assessments. Examples of diagnostic reading instruments are the *Stanford Diagnostic Reading Test, Fourth Edition* and the *Nelson-Denny Reading Test*. In the mathematics areas, counselors may see a report that used the KeyMath—3 or the *Sequential Assessment of Mathematics Inventories*.

Aptitude Assessment

Because this chapter addresses both achievement and aptitude assessment, it is important to begin a discussion on aptitude assessment that emphasizes that the purposes of aptitude assessments are quite different from achievement testing. Achievement testing provides information on what the individual has learned to-date, whereas aptitude testing provides an indication of how the individual will perform in the future. Clients often seek counseling to make decisions about the future; thus, knowledge of aptitude testing is often useful. Some aptitude instruments predict the ability to acquire a specific type of skill or knowledge, and others predict performance in a certain job or career. Aptitude instruments are frequently used to predict future academic performance (e.g., success in college). The pertinent issue with any aptitude instrument is the degree to which it predicts relative to the criterion of interest.

LO6 Scholastic Aptitude Tests

Hood and Johnson (2007) argued that counselors in a variety of settings (e.g., mental health agencies, elementary schools) need to be knowledgeable about the predominant scholastic aptitude tests. They contended that counselors in most settings are consulted by friends, relatives, and colleagues whose children are applying to colleges and universities. Traditionally, the public has expected counselors to be knowledgeable of these prevalent instruments. Hence, counselors who do not have at least a working knowledge of these instruments may not be viewed as credible by some individuals.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

SAT. The SAT is a well-established and widely used predictor of college performance. In 1994, the Scholastic Aptitude Test, which had been around since 1926, was revised and renamed the Scholastic Assessment Test. Part of the reason for changing the name of the SAT was to move away from any reference that might imply that the instrument measures innate aptitudes. The instrument originated to provide a common standard to which students could be compared because subject matter of high school courses and grading procedures varied so widely. The SAT is now known as simply the SAT.

Students can take the new and revised SAT starting March 2016. In October 2015, College Board will launch the redesigned PSAT/NMSQT and the new PSAT 8/9, which is for eighth- and ninth-graders. The new PSAT 10—for 10th-graders—will launch in February 2016. Tightly aligned, all exams include a reading test, a writing and language test, and a math test. This suite of assessments is designed to measure college and career readiness.

According to College Board, the SAT is the nation's most widely used college admission test. Although the SAT is offered throughout the school year; students frequently take it in the spring of 11th grade and again in the fall of 12th grade. Although the new and revised SAT has not been released at the writing of this book, I will focus on it since its publication corresponds to the publication of this book. Table 10.1 summarizes some of the major differences between the previous edition of the SAT and the new and redesigned SAT.

The new Evidence-Based Reading and Writing sections of the redesigned SAT will be different from previous version of the SAT in that there will be more focus on the students' abilities to interpret, synthesize, and use evidence found in a wide range of sources. These sources include

TABLE 10.1
Comparison of previous SAT with redesigned SAT

	Previous SAT	Redesigned SAT
Components	■ Critical Reading ■ Mathematics ■ Writing	■ Evidence-Based Reading and Writing ■ Reading Test ■ Writing and Language Test ■ Math ■ Essay (optional)
Important Features	■ Emphasis on general reasoning ■ Emphasis on vocabulary words often in limited context ■ Complex scoring (a point for a correct answer and a deduction for an incorrect answer)	■ Continued emphasis on general reasoning plus a focus on the knowledge, skills, and understandings important for college and career readiness and success ■ More emphasis on the meaning of words in contexts and on how word choice shapes meaning, tone, and impact ■ Rights-only scoring (a point for a correct answer but no deduction for an incorrect answer)
Essay	■ Required and given at start of SAT	■ Optional and given at end of SAT
Scoring	■ Scale ranging from 600 to 2400 ■ Scale ranging from 200 to 800 for Critical Reading; 200 to 800 for Mathematics; 200 to 800 for Writing	■ Scale ranging from 400 to 1600 ■ Scale ranging from 200 to 800 for Evidence-Based Reading and Writing; 200 to 800 for Math ■ Essay results reported separately
Subscore Reporting	■ None	■ Subscores are provided for added information

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

informational graphics, such as tables, charts, and graphs, as well as multiparagraph passages in the areas of literature, nonfiction, the humanities, science, social studies, and history. The Reading Test of the PSAT 8/9, PSAT 10, PSAT/ NMSQT, and the SAT becomes more complex as students advance in grade and the SAT requires more reading problems, longer reading passages, interpretation of more informational graphics, and greater emphasis on analysis of topics in history/social studies and science. The Writing and Language Test questions are multiple choice and are based on reading passages that may involve informational graphics. The Writing and Language Test ask students to make choices in regards to vocabulary, sentence structure, organization, tone, and factual support. Like the Reading Test, test items become more sophisticated as one progress from the PSAT 8/9 to the SAT.

As students progress from the PSAT 8/9, PSAT 10, PSAT/NMSQT to the SAT, the Math portion also increases in terms of the sophistication of the knowledge assessed. In one section of the Math portion students can use calculators, whereas in a second section they cannot. Later exams include more student-produced response questions (i.e., grid-ins) and fewer multiple-choice items. Also later exams include more complicated concepts and equations and student will need to complete multistep math problems. It should also be noted that these math problems are grounded in science, social science, career scenarios, and other real-life contexts.

With the redesigned SAT, the Essay portion of the exam is optional. Some colleges and universities, however, may still require this portion of the SAT. The Essay portion of the SAT requires students to demonstrate college and career readiness in reading, writing, and analysis. Students respond to a "high-quality source text" in which they must produce a clearly written and cogent analysis of the text material. Therefore, the Essay requires critical reasoning, knowledge of grammar, writing skill, and ability to draw evidence from the text.

It is anticipated that the scoring of the designed SAT will be similar to the previous editions of the SAT. The Evidence-Based Reading and Writing and the Math Tests each have scores from 200 to 800, which means the total SAT scores now vary from 400 to 1600. In the previous edition of the SAT, the Essay test was required so scores varied from 600 to 2400. As of the writing of this book, it is unclear how the optional Essay section of the redesigned SAT will be scored.

For school counselors, there is SAT Counselor Resource Center, which includes tools and information counselors may need to help students and parents through the college planning process. Educational Testing Service (ETS) uses a complex equating process that adjusts for variation in difficulty from one edition to the next. In addition, counselors can find significant free resources on the College Board website (http://www.collegeboard.org) to assist students in preparing for the SAT.

Before leaving our discussion of the SAT, we should also discuss SAT Subject Tests. The SAT Subject Tests are designed to measure students' knowledge in a specific area and are used for placement purposes in college or to test out of college coursework. These assess not just content knowledge but also whether students can apply the concepts pertinent to the particular areas. Students can take assessments in the areas of English, mathematics, science, and foreign languages. In the foreign languages tests, the students both listen and read in the tested language. Students are required to bring an acceptable CD player with earphones.

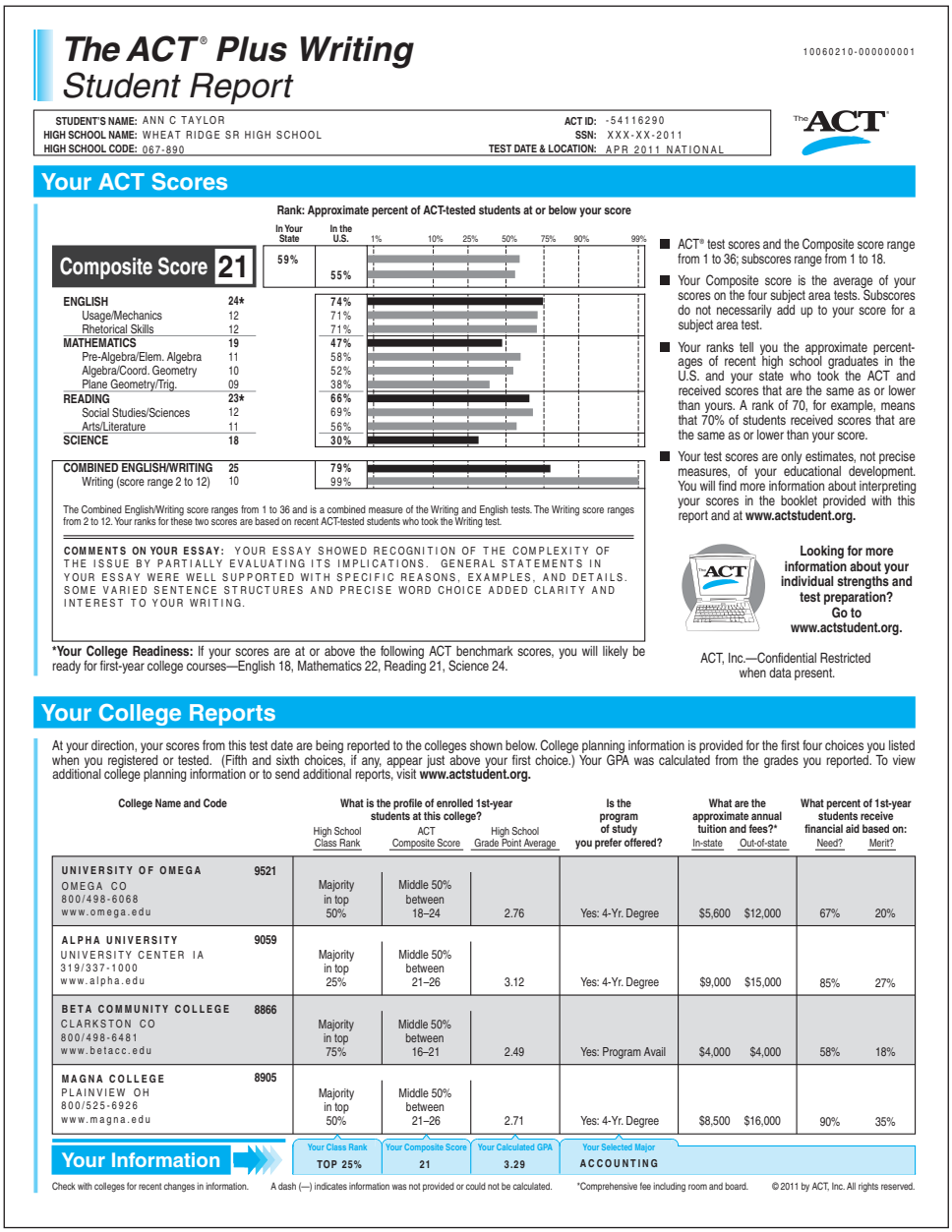
ACT Inc. The other major nationally administered college admissions testing program is the ACT. The ACT is the test students often take in either 11th or 12th grade, and many colleges use these scores in making decisions about admission. The ACT is a standards-based educational and career planning tool that assesses students' academic readiness for college. The ACT is the capstone of our College and Career Readiness System, which includes ACT Aspire that was discussed previously.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The developers of the ACT have attempted to measure as directly as possible the knowledge and skills students will need to perform college-level work. The content of the ACT is curriculum-based and tests students' knowledge in English, mathematics, reading, and science. There is an optional writing subtest.

An example of the ACT *Plus Writing Student Report* is included in Figure 10.1. As can be seen from scores in the upper left portion of the front of the report, students' performance is reflected in a Composite Score and the four major areas of English, mathematics, reading, and science. On these scales the scores range from 1 to 36 and percentile scores are included in the report. For the Composite Score, both state and national percentile

FIGURE 10.1
A sample of ACT results. Copyright © 2011 by ACT, Inc. All rights reserved. Reproduced with permission of ACT, Inc.



The ACT® Plus Writing Student Report

10060210-000000001

STUDENT'S NAME: ANN C TAYLOR
HIGH SCHOOL NAME: WHEAT RIDGE SR HIGH SCHOOL
HIGH SCHOOL CODE: 067-890
ACT ID: -54116290
SSN: XXX-XX-2011
TEST DATE & LOCATION: APR 2011 NATIONAL

The ACT

Your ACT Scores

Rank: Approximate percent of ACT-tested students at or below your score

		In Your State	In the U.S.
Composite Score	**21**	59%	55%
ENGLISH	24*		74%
Usage/Mechanics	12		71%
Rhetorical Skills	12		71%
MATHEMATICS	19		47%
Pre-Algebra/Elem. Algebra	11		58%
Algebra/Coord. Geometry	10		52%
Plane Geometry/Trig.	09		38%
READING	23*		66%
Social Studies/Sciences	12		69%
Arts/Literature	11		56%
SCIENCE	18		30%
COMBINED ENGLISH/WRITING	25		79%
Writing (score range 2 to 12)	10		99%

The Combined English/Writing score ranges from 1 to 36 and is a combined measure of the Writing and English tests. The Writing score ranges from 2 to 12. Your ranks for these two scores are based on recent ACT-tested students who took the Writing test.

COMMENTS ON YOUR ESSAY: YOUR ESSAY SHOWED RECOGNITION OF THE COMPLEXITY OF THE ISSUE BY PARTIALLY EVALUATING ITS IMPLICATIONS. GENERAL STATEMENTS IN YOUR ESSAY WERE WELL SUPPORTED WITH SPECIFIC REASONS, EXAMPLES, AND DETAILS. SOME VARIED SENTENCE STRUCTURES AND PRECISE WORD CHOICE ADDED CLARITY AND INTEREST TO YOUR WRITING.

***Your College Readiness:** If your scores are at or above the following ACT benchmark scores, you will likely be ready for first-year college courses—English 18, Mathematics 22, Reading 21, Science 24.

- ACT® test scores and the Composite score range from 1 to 36; subscores range from 1 to 18.
- Your Composite score is the average of your scores on the four subject area tests. Subscores do not necessarily add up to your score for a subject area test.
- Your ranks tell you the approximate percentages of recent high school graduates in the U.S. and your state who took the ACT and received scores that are the same as or lower than yours. A rank of 70, for example, means that 70% of students received scores that are the same as or lower than your score.
- Your test scores are only estimates, not precise measures, of your educational development. You will find more information about interpreting your scores in the booklet provided with this report and at **www.actstudent.org.**

Looking for more information about your individual strengths and test preparation? Go to www.actstudent.org.

ACT, Inc.—Confidential Restricted when data present.

Your College Reports

At your direction, your scores from this test date are being reported to the colleges shown below. College planning information is provided for the first four choices you listed when you registered or tested. (Fifth and sixth choices, if any, appear just above your first choice.) Your GPA was calculated from the grades you reported. To view additional college planning information or to send additional reports, visit **www.actstudent.org.**

College Name and Code	What is the profile of enrolled 1st-year students at this college? High School Class Rank	ACT Composite Score	High School Grade Point Average	Is the program of study you prefer offered?	What are the approximate annual tuition and fees?* In-state	Out-of-state	What percent of 1st-year students receive financial aid based on: Need?	Merit?
UNIVERSITY OF OMEGA 9521 OMEGA CO 800/498-6068 www.omega.edu	Majority in top 50%	Middle 50% between 18–24	2.76	Yes: 4-Yr. Degree	$5,600	$12,000	67%	20%
ALPHA UNIVERSITY 9059 UNIVERSITY CENTER IA 319/337-1000 www.alpha.edu	Majority in top 25%	Middle 50% between 21–26	3.12	Yes: 4-Yr. Degree	$9,000	$15,000	85%	27%
BETA COMMUNITY COLLEGE 8866 CLARKSTON CO 800/498-6481 www.betacc.edu	Majority in top 75%	Middle 50% between 16–21	2.49	Yes: Program Avail	$4,000	$4,000	58%	18%
MAGNA COLLEGE 8905 PLAINVIEW OH 800/525-6926 www.magna.edu	Majority in top 50%	Middle 50% between 21–26	2.71	Yes: 4-Yr. Degree	$8,500	$16,000	90%	35%

Your Information

Your Class Rank	Your Composite Score	Your Calculated GPA	Your Selected Major
TOP 25%	21	3.29	ACCOUNTING

Check with colleges for recent changes in information. A dash (—) indicates information was not provided or could not be calculated. *Comprehensive fee including room and board. © 2011 by ACT, Inc. All rights reserved.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 10.1
(continued)

ANN C TAYLOR

Planning Your Education and Career

Many people consider several possibilities before making definite career plans. Before you took the ACT®, you had the opportunity to respond to questions about your educational and career plans. Use this information to consider possibilities that you may like to explore.

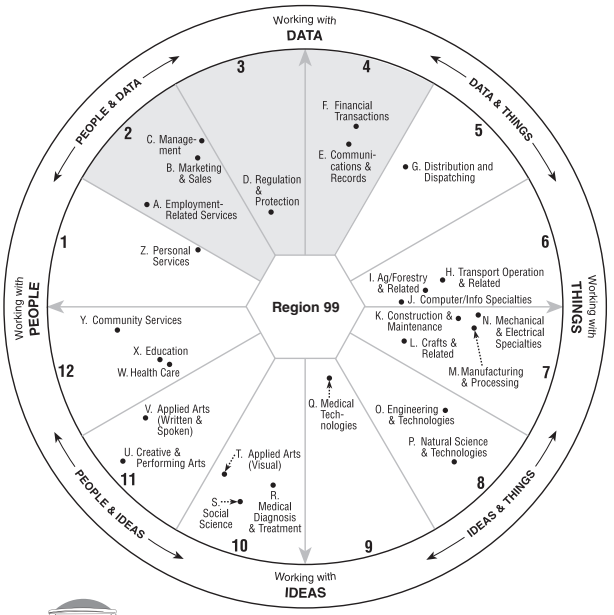
Your Interest Inventory Results	The College Major You Indicated	The Occupational Field You Indicated
YOUR RESULTS INDICATE A PREFERENCE FOR WORKING WITH PEOPLE AND DATA.	ACCOUNTING	INSURANCE & RISK MANAGEMENT
SEE MAP REGIONS 2, 3, 4 THE SHADED REGIONS SHOW CAREER AREAS HAVING WORK TASKS YOU PREFER.	THIS MAJOR PRIMARILY INVOLVES WORKING WITH DATA AND THINGS.	**SEE MAP REGION 2** THE OCCUPATIONAL FIELD YOU CHOSE IS IN CAREER AREA C: MANAGEMENT
RELATED CAREER AREAS: COMMUNICATIONS & RECORDS EMPLOYMENT-RELATED SERVICES FINANCIAL TRANSACTIONS MANAGEMENT MARKETING & SALES REGULATION & PROTECTION	**RELATED MAJORS:** BANKING & FINANCIAL SUPPORT SERVICES BUSINESS ADMINISTRATION & MGMT, GEN FINANCE, GENERAL FINANCIAL PLANNING & SERVICES INSURANCE & RISK MANAGEMENT INVESTMENTS & SECURITIES PURCHASING/PROCUREMENT/CONTRACTS MGMT SMALL BUSINESS MANAGEMENT/OPERATIONS	**RELATED OCCUPATIONS:** ASSOCIATION EXECUTIVE FINANCIAL MANAGER FOREIGN SERVICE OFFICER GENERAL MANAGER/TOP EXECUTIVE HOTEL/MOTEL MANAGER MANAGEMENT CONSULTANT PROPERTY/REAL ESTATE MANAGER

Your Guide to College and Career Planning

- All college majors and occupations differ in how much they involve working with four basic work tasks: working with **People** (care, services), **Things** (machines, materials), **Data** (facts, records), and **Ideas** (theories, insights). These four basic work tasks are the compass points on the World-of-Work Map.
- The map is divided into 12 regions, each with a different mix of work tasks. The map shows the locations of 26 occupational fields, called Career Areas (A-Z). Each Career Area contains many occupations that share similar work tasks.

*If no regions are shaded, you did not answer enough interest items to permit scoring.

The World-of-Work Map
(Your Interest Inventory results are shaded.*)



ACT

For more information about your college and career planning, visit **www.actstudent.org** or check the booklet provided with this report.

16627

scores are provided. There are also subscale scores in English, Mathematics, and Reading, which are designed to provide more detailed information about strengths and limitations in those areas. The subscale scores range from 1 to 18, and percentiles are reported for each of the subscales. If students take the optional Writing Test, which is often required for college entrance and/or course placement, then their report will include an English/Writing composite score using the same range of 1–36. In addition, there is a Writing subscale score and some comments related to the student's essay. In addition, ACT (2014) provided benchmark scores for each of the four major academic areas in regard to the chances of getting a B or C in related coursework at the college level. If a score is marked with an asterisk (*), then there is a 50% chance of obtaining a B or higher or about a 75% chance of obtaining a C or higher in the corresponding first-year college course.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Some counselors will primarily focus on the ACT academic test scores. The ACT report, however, includes other pertinent information that can provide admissions-related information regarding institutions the student ranked as his or her top four choices. This information enables students and parents to easily see how a particular student's performance compares with those who have been accepted at that institution. The reverse side of the ACT report addresses planning the student's education and career. While students are taking the ACT, they also take an interest inventory. The results of the interest inventory are summarized in the "Planning Your Education and Career" section of the report. This section also includes a list of college majors and occupational fields that correspond to the student's interests. The World of Work Map also provides a visual representation of the student's interests and can be used by counselors to assist students in exploring careers that are related to their particular interests. The World of Work schemata has been extensively researched and represents students' preferences regarding working with data, ideas, people, and things. Within the map are occupations so that students can easily identify occupations to explore.

Graduate Record Examination. Another scholastic aptitude test is the Graduate Record Examination (GRE). Unlike the other assessments, the GRE is designed to predict performance in graduate school as compared with undergraduate college performance. The GRE General Test is a generic test for prospective graduate students that was revised starting in August 2011. The GRE revised General Test has new questions that were designed to more accurately reflect graduate school skills. Furthermore, as it is only given by computer in the United States, the new revised test lets examinees now change answers and skip questions. Many academic programs want to recruit students who already have knowledge about the discipline, and to do so they use one of the GRE subject tests to assess this knowledge. Other tests that have been developed in specific disciplines for the selection of students into professional schools include the Medical College Admissions Test (MCAT), the Dental Admission Test (DAT), the Law School Admission Test (LSAT), and the Graduate Management Admission Test (GMAT).

This discussion covers only the GRE revised General Test, which will be referred to as the GRE. The GRE consists of three sections: Verbal Reasoning, Quantitative Reasoning, and Analytical Writing. The scores for the Verbal Reasoning and Quantitative Reasoning range in 1-point increments from 130 to 170, whereas the Analytical Writing will continue to range from 0 to 6. The Verbal Reasoning now has no analogies or antonyms and there is more of a focus on reading. Quantitative Reasoning includes basic mathematical concepts (arithmetic, algebra, geometry, and data analysis) with more emphasis on data interpretation. As a part of the testing process, there is an onscreen calculator. The Analytical Writing section is designed to measure how effectively an individual can articulate complex ideas, evaluate accompanying evidence, support ideas appropriately, provide a coherent discussion, and control the elements of standard written English. An individual's essay is evaluated by two trained readers using a 6-point rating scale.

The GRE is offered as a computer-based assessment in the United States, Canada, and many other countries. The paper-based version is available in countries where computer-based testing is not available. There are numerous testing sites throughout the United States. Individuals can schedule to take the GRE at a time convenient to them and, therefore, do not need to wait until a scheduled testing date. The GRE was one of the first assessments to take advantage of computer-adaptive assessment; in other words, the computer adapts the test to the individual taking the GRE. At the beginning of the test, the test taker is presented with questions requiring average ability. As the individual answers questions, the computer analyzes the answers to determine the difficulty of the next question. The scoring of the test takes into consideration the number of correct responses and the difficulty of the items.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Validity of Scholastic Aptitude Tests. The most pertinent questions about scholastic aptitude tests relate to how well they predict collegiate performance. These tests are used to make decisions that often have a significant influence on people's lives; therefore, the validity of these instruments deserves analysis. In general, research indicates that the validity of the SAT and the ACT are about the same. Many of the studies indicate that the correlation between scores on these tests and freshman grade point average (GPA) is in the .30 to .50 range. Korbin, Patterson, Shaw, Mattern, and Barbuti (2008) found that of the three SAT scores, the Writing scale had the largest correlation ($r = .33$) with freshman-year GPA. The best predictor, however, is a combination of high school grades and all three SAT section scores. These two measures seem to balance each other—grades take into consideration persistence and hard work, whereas the scholastic aptitudes tests are not influenced by partiality or teacher bias.

The same trend applies to predicting graduate school performance—multiple predictors are better than any one single predictor. In a meta-analysis of 1,753 independent correlations, Kuncel, Hezlett, and Ones (2001) found that the relationship between GRE scores and first-year graduate GPA, comprehensive exam scores, and faculty ratings were significant, and the average correlational coefficients ranged from .26 to .47. They also found across most disciplines that the combination of GRE scores and undergraduate GPA tended to predict better than GRE scores alone. Although many of the studies on the GRE have assessed students' GPAs during their first year in graduate school, Burton and Wang (2005) found that GRE scores were significantly related to cumulative GPA. Consistent with other researchers, Burton and Wang found that combining GRE scores with undergraduate GPA was better at predicting cumulative GPA. The Graduate Record Examination Board (2014) also recommended using multiple sources of information (e.g., GPA, letters of recommendation) in making admission decisions.

Institutions vary in the degree to which scholastic aptitude test scores influence admissions decisions. To counsel college-bound individuals effectively, counselors need to have knowledge about admissions requirements and typical test score performance. For counselors working with scholastic aptitude tests, it is important to remember these instruments do provide some information but that information should be evaluated in a broader context. Publishers of scholastic aptitude assessments suggest that scores from these instruments be combined with other information. Counselors need to help clients sort out the meaning of their scores, a task that can be facilitated by incorporating other germane information into the interpretation of scholastic aptitude assessments.

LO7 Vocational/Career Aptitude Tests

As mentioned earlier, aptitude tests are often used to predict academic or vocational/career performance. In addition to the scholastic aptitude tests we have discussed, numerous vocational aptitude tests are used in career counseling to predict job or occupational performance. Although these vocational or career aptitude instruments are discussed here rather than in the chapter on career assessment, they should also be considered when selecting assessment tools in career counseling.

There are also vocational aptitude assessments in which the primary purpose is not for use in counseling but rather for employment selection purposes. Effective selection requires that an instrument accurately predict successful performance of job duties. This is a difficult task because job duties within an occupation vary depending on the organization and the setting. In addition, if the instrument is used to predict performance in an occupation, then the test needs to be validated with information from people who are performing successfully in that occupation. This necessitates a precise definition of what constitutes performing "successfully" in that occupation—simply being employed in a job

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

does not guarantee success or competence. This matter is further complicated by the fact that job performance can rarely be measured in a unidimensional manner. Most occupations are complex, and some authorities in the field contend that multifaceted methods of evaluating job performance must be utilized (Campbell, 1994). Therefore, all of these factors make validating an instrument designed to predict job performance a difficult and time-consuming task. The difficulty in gathering information on job performance increases the difficulty in getting large norming groups. This problem is magnified when test developers want to create an instrument that predicts performance for a number of occupations. A test developer needs a sufficient number of individuals successfully performing in each occupation to have an adequate norming group.

Some vocational aptitude tests are designed to measure one aptitude or to predict performance in a single occupational area. For example, psychomotor tests are single-purpose aptitude tests that were originally thought to be predictive of many occupations. These instruments are predictive of lower-level manual skills, but they are not good predictors of higher-order abilities (Anastasi, 1988). Clerical aptitude and mechanical aptitude tests are examples of single-purpose vocational aptitude tests that can predict success in certain occupations. However, clients frequently want to know their strengths and limitations and how they might perform in various occupations. Because most clients do not want to spend time taking numerous single-purpose aptitude tests, aptitude *batteries* can be helpful. Aptitude batteries assess multiple aptitudes and predict performance in various occupations. These aptitude batteries can assist clients in examining their strengths and limitations, and also assess potential in a variety of career areas.

Armed Services Vocational Aptitude Battery. The *Armed Services Vocational Aptitude Battery* (ASVAB), a part of the ASVAB Career Exploration Program, is the most widely used aptitude test in the United States. The Department of Defense offers the ASVAB Career Exploration Program, free of charge, to students in secondary and some postsecondary schools. The ASVAB Career Exploration Program consulted with a number of noted researchers in the field of vocational psychology and produced a comprehensive career program that is competency-based. The redesigning of the ASVAB Career Exploration Program began in 2001, and this new program provides students with a more comprehensive career exploration and planning program (U.S. Department of Defense, 2011). The program is targeted for all students in the 10th, 11th, and 12th grades along with individuals in postsecondary schools. Web-based and printed materials are designed to assist students in exploring their results and possible career options. There are also structured activities that counselors can use to facilitate the career exploration program.

The Career Exploration Program begins with students taking the ASVAB. Typically a school counselor will arrange for a post-session that focuses on the students' ASVAB scores. This session, typically conducted by a Department of Defense (DoD) education service specialist, also includes administering, scoring, and interpreting the Find Your Interest (FYI) inventory. The FYI is based on Holland's model of career interests. The students also learn how to use *Exploring Careers: The ASVAB Career Exploration Guide.* In addition, the students are shown how to use their results with OCCU-Find, which assists students in identifying occupations that correspond to their interest. Counselors can also facilitate a comprehensive career exploration program—additional materials supplement these activities. More details can be obtained from the ASVAB website at http://www.asvabprogram.com.

Ryan Krane and Tirre (2005) contended that the ASVAB itself is distinguished by its superior norms, thorough investigation of fairness, and substantial evidence of criterion-related validity. The ASVAB is composed of 200 items and requires approximately 3 hours to complete. Students take eight subtests: General Science, Arithmetic Reasoning,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Word Knowledge, Paragraph Comprehension, Mathematics Knowledge, Electronics Information, Auto and Shop Information, and Mechanical Comprehension. There are three composite scores (see Figure 10.2: Verbal Skills, Math Skills, and Science and Technical Skills). These Career Exploration Scores are designed to provide students with a general sense of their skills in the three areas as compared with other students in the same grade. These composite skill scores can also be easily used to search for occupations, which is a somewhat unique factor of the ASVAB Career Exploration Program. As seen in Figure 10.2 the ASVAB Summary Results provides results using T scores with corresponding error bands. The percentile scores are reported for each sex separately and then for a combined norm. The Military Entrance Score (also known as the AFQT) indicates whether a student has scored high enough to meet different entrance requirements within the military.

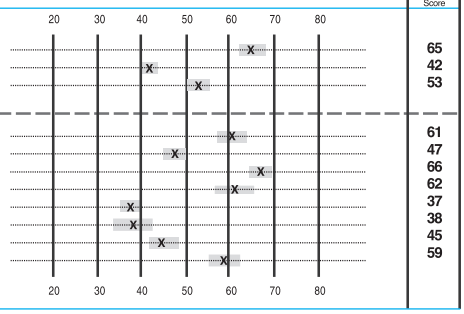
The national sample for the ASVAB was obtained from aptitude test data collected as a portion of the Profile of American Youth (PAY97) project. Two samples were drawn for it: (a) Enlisted Training Program and (b) Career Exploration Program. The Enlisted Training Program sample involved 6,000 individuals and is used to screen youth for services in the military. The Career Exploration Programs involved 4,700 students in high school and provides the norms for students taking the ASVAB in grades 10th, 11th, and 12th. The reliability estimates were derived based on Item Response Theory, and the coefficients for the composite

FIGURE 10.2
ASVAB Summary Results Sheet

Student
12th Gr Female (Form 23G)
SSN: XXX-XX-9999
Test Date: Jul 11, 2006
Old Dominion H.S.
Hometown DC
Print No.:XXXXX

ASVAB SUMMARY RESULTS

ASVAB Results	Percentile Scores 12th Grade Females	12th Grade Males	12th Grade Students	12th Grade Standard Score
Career Exploration Scores				
Verbal Skills	97	95	96	65
Math Skills	22	17	19	42
Science and Technical Skills	81	48	64	53
ASVAB Tests				
General Science	91	81	86	61
Arithmetic Reasoning	43	30	37	47
Word Knowledge	98	95	96	66
Paragraph Comprehension	92	91	91	62
Mathematics Knowledge	14	12	13	37
Electronics Information	13	10	11	38
Auto and Shop Information	53	21	37	45
Mechanical Comprehension	95	76	85	59
Military Entrance Score (AFQT) 57				



EXPLANATION OF YOUR ASVAB PERCENTILE SCORES

Your ASVAB results are reported as percentile scores in the three highlighted columns to the left of the graph. Percentile scores show how you compare to other students - males and females, and for all students - in your grade. For example, a percentile score of 65 for an 11th grade female would mean she scored the same or better than 65 out of every 100 females in the 11th grade.

For purposes of career planning, knowing your relative standing in these comparison groups is important. Being male or female does not limit your career or educational choices. There are noticeable differences in how men and women score in some areas. Viewing your scores in light of your relative standing both to men and women may encourage you to explore areas that you might otherwise overlook.

You can use the **Career Exploration Scores** to evaluate your knowledge and skills in three general areas (Verbal, Math, and Science and Technical Skills). You can use the **ASVAB Test Scores** to gather information on specific skill areas. *Together, these scores provide a snapshot of your current knowledge and skills.* This information will help you develop and review your career goals and plans.

EXPLANATION OF YOUR ASVAB STANDARD SCORES

Your ASVAB results are reported as standard scores in the above graph. Your score on each test is identified by the "X" in the corresponding bar graph. You should view these scores as *estimates* of your true skill level in that area. If you took the test again, you probably would receive a somewhat different score. Many things, such as how you were feeling during testing, contribute to this difference. This difference is shown with gray score bands in the graph of your results. Your standard scores are based on the ASVAB tests and composites based on your grade level.

The score bands provide a way to identify some of your strengths. Overlapping score bands mean your true skill level is similar in both areas, so the real difference between specific scores might not be meaningful. If the score bands do not overlap, you probably are stronger in the area that has the higher score band.

The ASVAB is an aptitude test. It is neither an absolute measure of your skills and abilities nor a perfect predictor of your success or failure. A high score does not guarantee success, and a low score does not guarantee failure, in a future educational program or occupation. For example, if you have never worked with shop equipment or cars, you may not be familiar with the terms and concepts assessed by the Auto and Shop Information test. Taking a course or obtaining a part-time job in this area would increase your knowledge and improve your score if you were to take it again.

USING ASVAB RESULTS IN CAREER EXPLORATION

Your career and educational plans may change over time as you gain more experience and learn more about your interests. *Exploring Careers: The ASVAB Career Exploration Guide* can help you learn more about yourself and the world of work, to identify and explore potential goals, and develop an effective strategy to realize your goals. The *Guide* will help you identify occupations in line with your interests and skills. As you explore potentially satisfying careers, you will develop your career exploration and planning skills.

Meanwhile, your ASVAB results can help you in making well-informed choices about future high school courses.

We encourage you to discuss your ASVAB results with a teacher, counselor, parent, family member or other interested adult. These individuals can help you to view your ASVAB results in light of other important information, such as your interests, school grades, motivation, and personal goals.

USE OF INFORMATION

Personal identity information (name, social security number, street address, and telephone number) and test scores will not be released to any agency outside of the Department of Defense (DoD), the Armed Forces, the Coast Guard, and your school. Your school or local school system can determine any further release of information. The DoD will use your scores for recruiting and research purposes for up to two years. After that the information will be used by the DoD for research purposes only.

MILITARY ENTRANCE SCORES

The **Military Entrance Score** (also called AFQT, which stands for the Armed Forces Qualification Test) is the score used to determine your qualifications for entry into any branch of the United States Armed Forces or the Coast Guard. The **Military Entrance Score** predicts in a general way how well you might do in training and on the job in military occupations. Your score reflects your standing compared to American men and women 18 to 23 years of age.

Use Access Code: 123456789X
(for online Occu-Find and FYI)
Access code expires: Jul 15, 2007
Explore career possibilities by using your Access Code at
www.asvabprogram.com

SEE YOUR COUNSELOR FOR FURTHER INFORMATION

DD FORM 1304-5, JUL 05 - PREVIOUS EDITIONS OF THIS FORM ARE OBSOLETE

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

scores on Verbal Skills, Math Skills, and Science and Technical Skills ranged from .84 to .92 (U.S. Department of Defense, 2011). A copy of a technical chapter related to the ASVAB is available at http://www.asvabprogram.com. Patrick, Blosel, and Gross (2009) contended that the ASVAB Career Exploration Program can help students learn more about themselves through the interest inventory (FYI) and the aptitude test (ASVAB) and explore occupations within both the civilian workforce and the military. They rated the materials to be of exceptional quality. The ASVAB is evolving, and numerous studies of job performance are continuing to be performed on this widely used instrument. In addition, the Department of Defense is in the process of developing a computer-adapted version that will shorten the time required for taking the ASVAB.

WorkKeys. WorkKeys is a *job skills* assessment system that is primarily used in the selection, hiring, training, and retaining of the workforce. Although I am discussing it related to vocational aptitude, it is a complex system that can be used by educators, counselors, or employers. I will focus primarily on how it can be used by counselors and educators. The ACT has one of the largest occupational profiles databases available, which includes more than 20,000 occupations that have been extensively studied to identify the essential skills and skill levels for employee selection, hiring, and training. It can be taken by individuals to get a clear picture of their skills and the skill levels needed for multiple occupations.

The WorkKeys Assessment assesses both cognitive and noncognitive skills. The cognitive assessments are Applied Mathematics, Applied Technology, Business Writing, Listening for Understanding, Locating Information, Reading for Information, Teamwork, and Workplace Observation. The noncognitive assessments are Fit, Performance, and Talent. Most educational settings have students take three scales: Applied Mathematics, Reading for Information, and Locating Information tests. If students do well enough (i.e., scores of 3) on each of these three scales, then they qualify for the *National Career Readiness Certificate*. The National Career Readiness Certificate is issued by ACT and is recognized by employers nationally as an indicator of readiness for entry-level positions.

Differential Aptitude Test. The last test to be discussed as an example of aptitude assessments is the *Differential Aptitude* Test (DAT; Bennett, Seashore, & Wesman, 1990). Although the DAT was frequently used in the past, its use has recently decreased. The DAT is designed for students in grades 7 through 12 but can also be used with adults. This instrument can be used alone or in conjunction with the Career Interest Inventory. A computer-adaptive version of the DAT takes about 90 minutes, whereas the paper- and-pencil version takes about 3 hours. The aptitudes measured, which are similar to those measured by many other instruments, include verbal reasoning, numerical reasoning, abstract reasoning, mechanical reasoning, space relations, spelling, language usage, and clerical speed and accuracy. The verbal reasoning and numerical reasoning scores are combined to produce the Scholastic Aptitude score. The DAT uses separate norms for males and females. The use of separate gender norms is related to findings that males and females tend to score differently on a few of the DAT scales. The norming groups for the DAT are large and were stratified to match the 1980 census information. The reliability coefficients are in the .80 to .90 range (Psychological Corporation, 1991). If results are computer-scored rather than hand-scored, the scores appear with confidence bands.

Critics of the Differential Aptitude Tests focus on the instrument's validity. Willson and Stone (1994) pointed out that predictive validity information is conspicuously absent. In addition, there is very little differential validity evidence. The term used in the composite score, *scholastic aptitude*, is important because the DAT has been found to be highly correlated with achievement tests and high school and college GPA (Psychological

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Corporation, 1991). A counselor can use the DAT to predict academic performance but should be very cautious in using the instrument to predict vocational or occupational performance. In addition, Wang (2002) noted the need for norms for ethnic minorities and methods for using the DAT norming samples with clients of color.

LO8 Test Preparation and Performance

Before concluding our discussion of achievement and aptitude testing, it is important to examine briefly the research related to the effects of test preparation. You have probably seen advertisements selling a set of workbooks or a CD-ROM or a special workshop guaranteed to give individuals higher scores on certain tests (e.g., the SAT or the MCAT). You may have questioned whether these programs or products actually improve scores significantly. The answer to that question is that it depends on the program. Before exploring which types of training programs have an impact on scores, some terms need to be defined. **Test sophistication** is a term applied to an individual's level of knowledge in test-taking skills. Test sophistication is not related to knowledge of the content but rather to the format of the tests and the skills required for maneuvering through that format. A distinction also needs to be made between the terms **coaching** *or education*, which are both related to instruction in the content of the examination. *Coaching* involves training or practice on questions that are the same as or similar to the items on the test. The intent is to help the test taker do well on the test, not necessarily to learn the content. *Education* occurs when the domain or area is covered more broadly, with the intent of helping the test taker learn the content or information.

Test Sophistication and Training

The question here is, Do programs that provide practice on the testing format make a significant difference? Some research indicates that test scores do improve when individuals retake an alternate form of the test (Donlon, 1984). Furthermore, individuals with extensive experience in taking standardized tests have an advantage. However, short orientation and practice sessions tend to bring individuals without extensive experience equal to those who have had more exposure to tests (Anastasi, 1981). These findings indicate that those who can afford test sophistication training have an advantage over those who cannot afford it. In the past few years, the registration materials for a number of tests (e.g., SAT, GRE) have included booklets or webpages that address test-taking strategies, explain item formats, and include sample test questions. These booklets or webpages are designed to provide those who cannot afford test sophistication training the same sort of experiences as those who can, thus producing "a more level playing field." Also, some of the websites for many of the well-known instruments have practice tests or sample questions to acquaint the individual with test-taking strategies. In addition, some schools provide instruction related to test-taking skills so that all students can use that information.

Coaching

If a program involves coaching, then it involves more than just test sophistication skills. The effects of coaching have been investigated for many years. Some research findings have found that coaching programs can make a significant difference in scores. Some of this research, however, has methodological flaws, particularly related to the equivalence of the treatment and control groups (Bond, 1989). There can be significant differences in scores if a researcher simply compares those who participated in these expensive coaching programs with those who did not. There are other significant differences between these two groups. For example, the members of the group that can afford the coaching

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

program are probably more affluent and have access to other resources that may increase their scores (e.g., tutoring, travel experiences, trips to museums). In general, the studies on coaching have found mixed and inconsistent results concerning the effectiveness of these programs.

One consistent finding is that the closer the resemblance of the coaching material is to the actual test content, the greater the improvement in test scores. A "teaching to the test" approach is one in which the focus is on learning the correct answer to individual questions rather than teaching the broader domain that the test is attempting to measure. A study by Powers and Rock (1999) indicated that the effect of coaching on the SAT, above the expected gain, was only about 26 points, with the average gain being greater on the Mathematics section (18 points) as compared with the Verbal section (8 points). Dominque and Briggs (2009) found similar findings in that the effects of coaching were only roughly 11 to 15 points on the math section of the SAT and 6 to 9 points on the verbal section. There are some indications that longer coaching programs result in higher gains as compared with shorter programs. Once again, this makes sense in that the more time individuals spend reviewing the material, the more likely it is that they will cover material that will be on the test. The general conclusions are that coaching programs may increase scores slightly, but significant changes occur only if the programs are longer and the content is closely aligned with the material on the test. For many clients, reviewing material on their own can be as effective as these programs and significantly less expensive.

Summary

At some point in their professional lives, counselors can expect to be called upon to have knowledge of both achievement and aptitude assessment. The distinction between an achievement test and an aptitude test is that the former measures past learning whereas the latter predicts potential ability to learn or perform.

In order to understand current trends in achievement testing, counselors need to understand the influence of No Child Left Behind. Although each state is now granted a waiver from the original No Child Left Behind legislation, there is still a need to show that all students are college and career ready, which requires the use of achievement assessments. In addition, the Common Core State Standards are also influencing achievement testing in many states. One example of an achievement test designed to correspond to the Common Core is Smarter Balanced.

In understanding achievement testing, counselor should also recognize the importance of individual achievement tests. Examples of frequently used individual achievement tests are the Kaufman Test of Educational Achievement, Third Edition (KTEA-3) and the Wechsler Individual Achievement Test, Third Edition (WIAT-III). Individual achievement tests are frequently used in the diagnosis of learning disabilities.

Aptitude assessment is performed to make predictions about the future. Often aptitude assessments predict educational or vocational/career performance. In the educational area, scholastic aptitude assessment frequently involves predicting performance in college, such as the SAT and ACT. Scholastic aptitude assessment instruments are often used for student selection and placement decisions. They can also serve as an effective counseling tool and assist clients in making decisions. There are also a number of vocational/career aptitude assessments. Professional counselors need to understand the strengths and limitations of any aptitude tool they use. It is particularly important to know how good a predictor the instrument is and what precisely is being predicted (e.g., freshman GPA, success in a training

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

program, job performance). If counselors do not understand the norming group and the validation information for an aptitude test, they may provide the client with misinformation that could be a disservice to that client.

Finally, counselors need to understand the difference between teaching test sophistication skills and coaching. Teaching test sophistication skills is usually brief and addresses test-taking skills, which tend to have minimal effect on individuals' performance on achievement and aptitude assessments. Coaching is typically more involved and includes test content. In general, the content and the amount of time spent in coaching activities tend to influence test performance.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 11

Assessment in Career Counseling

Learning Objectives

- **LO1** Understand the importance of assessment of interests in career counseling and describe major interest assessments
- **LO2** Describe the reasons and methods for assessing skills and abilities in career counseling
- **LO3** Identify the reasons and methods for assessing values in career counseling
- **LO4** Describe three types of career process measures
- **LO5** Identify different types of qualitative career assessments
- **LO6** Distinguish between the advantages and disadvantages of technology in career assessment
- **LO7** Understand gender issues in career assessment
- **LO8** Explain issues related to race, ethnicity, or culture in career assessment

Enrico is a 24-year-old Latino male whose parents emigrated from Argentina to the United States when he was 3 years old. Enrico received mostly As and Bs in high school, and his parents strongly encouraged him to go to college and major in medicine. He followed their advice and was admitted to a large Midwestern university. Enrico, however, struggled with both biology and chemistry at the college level and changed his major to business after the first semester. In his second semester, Enrico took an accounting course, which he found boring, so he stopped attending class after a few weeks. He once again decided to change his major and selected Spanish, which was the language often spoken at his parents' home. Enrico's parents, however, were disappointed in his selection of Spanish as a major and encouraged him to return to either medicine or business. An advisor in the on-campus academic advising center suggested that Enrico seek career counseling to help him select a major and determine a career direction. Enrico, frustrated by his indecision about a major, sought career counseling with the hopes of finding a stable career direction.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

If you were Enrico's counselor, you might consider incorporating one or more career assessments into the counseling process. Although some individuals believe that career counseling always involves formal assessments, that is not always the case. The question is whether an assessment or assessments would assist Enrico and, further, how these assessments could be used in the counseling process to provide Enrico with pertinent information. Sometimes, potential clients like Enrico believe there exists a standard battery of career tests that are given to clients to provide definitive information on the "best" career for an individual. There is not, however, a standard battery of career assessments that is appropriate for all clients. Therefore, as Enrico's counselor, you must be informed on the array of assessments available so that you can make sound decisions about which assessments are most appropriate. Furthermore, using career assessments requires interpretative skills and clinical competence. The choice of appropriate career assessments is important because the selection of an occupational direction can affect whether individuals are employed or unemployed, whether they are successful in their occupational pursuits, and how satisfied they are with activities that typically involve at least eight hours a day.

Historically, assessment has been considered an integral part of career counseling. Career assessment has a long history, and many career counselors trace its origins back to Frank Parsons. Parsons (1909) encouraged career assessment in his three-step model of career counseling. In fact, the first step was to study or measure the individual, which often involved testing the client. Some people still view career counseling as solely consisting of testing and then providing occupational information. This "test and tell" approach is not reflective of the current status of this counseling area because research has shown that career counseling is not separate from personal counseling (Swanson, 2012). Assessment, however, is often involved in career counseling, and numerous instruments are routinely used to help individuals explore career directions and to make effective career decisions. Spokane (1991) proposed that the purposes of career assessments are to unearth career possibilities congruent with the client's attributes, to assess conflicts and problems, to motivate constructive behavior, to acquire a cognitive structure for evaluating career alternatives, to clarify expectations and plan interventions, and to establish the range of abilities. As reflected in Spokane's description, career assessment has diverse goals and covers a wide variety of areas. Although this chapter focuses on some commonly used career instruments, career assessment should not be viewed as a discrete set of instruments. The American Psychological Association (APA) Task Force on Test Users Qualification argued that test users in a career/vocational area must also be able to assess personality and mental health problems that may impede successful career development (Turner, DeMers, Fox, & Reed, 2001).

According to Schwiebert (2009) the purpose of any career assessment is to gather information to facilitate decision making by the counselor or client. This chapter divides career assessment into two major categories: tools used to assess individual differences and tools used to assess the career development process. The first category of career assessments is designed to measure different aspects of an individual, such as interests, abilities, values, and needs. There are also integrated assessments that involve various career assessments. These measures of individual differences are useful in career counseling because of their relationship to helping clients make effective career choices or decisions. The second category of instruments relates to the career process. The focus of these instruments is not on the individual attributes of a client but rather on where the client is in the process of selecting a career. In terms of the career process measures, many of these instruments assess either the client's level of indecision or the client's stage of career maturity or adaptability.

As mentioned earlier, not all career-related assessments need to be formal career instruments. Some clinicians have been influenced by postmodern philosophy and constructivism and believe qualitative career assessment can be useful either as replacement

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

for more formal assessment or in conjunction with published instruments. This chapter also includes a section on qualitative career assessment.

Additionally, this chapter addresses issues in career assessment. One of these issues is related to structuring inventories in a way that minimizes gender stereotypes. Research indicates that counselors need to interpret career measures with a sensitivity to sex-role socialization and to encourage clients to explore nontraditional occupations that may correspond to their attributes. There are also ethnic and racial issues related to career assessments. Racial differences on general ability and aptitude assessments must also be reexamined because many of the instruments discussed in previous chapters are used in career counseling. Furthermore, racial and ethnic issues are also pertinent to other types of career assessments used by counselors.

LO1 Assessing Individual Differences

Interests

Interest inventories are often used in career counseling because they can be helpful in describing an individual's general occupational interests. Yet, Sodano (2015) indicated that using interest inventories is not the only method for assessing interests. Counselors can also assess interests by using measures of expressed or manifest interests. Exploring *expressed interest* involves simply asking clients about their interests, which can be a very useful indicator of vocational interests, particularly with adults. *Manifest interests*, on the other hand, are those interests identified by examining how clients choose to spend their time and the activities they select. Assessing both expressed and manifest interests can be a good method for identifying vocational interests. In particular, research has found that expressed and measured interests are equally good at predicting occupational choice (Betsworth & Fouad, 1997). The relationship between expressed and measured interests is significant, with the average correlation being around .46 (Athanasou & Cooksey, 1993). Although these methods of assessing interests can provide sound information, there are some advantages to using more formal interest inventories. Interest inventories have been found to promote career exploration (Herr, Cramer, & Niles, 2004), and many interest inventories connect the client's interests to specific occupations. Some clients have a general idea about their interests but may not know which occupations are related to those interest areas.

In assessing interests, it is important for the counselor to understand the relationship between interests and certain career variables. For example, some clinicians believe that if clients are highly interested in a specific field, then those clients will be successful in that career area. Strong interests do not guarantee occupational success (Betsworth & Fouad, 1997). When explaining the results of an interest inventory, the counselor should stress that these instruments are measuring only interests and do not provide an indication of ability. In summarizing existing research on the relationship between interests and abilities, Lent, Brown, and Hackett (1994) reported a small but significant relationship between them. The relationships between interests and occupational satisfaction have also been found to be small (Kristof-Brown, Zimmerman, & Johnson, 2005). These small relationships probably result from the fact that work satisfaction involves multiple variables (e.g., monetary rewards, colleagues). Interests are, however, good predictors of career direction (Pryor & Taylor, 1986). Although interest inventories have been found to be good predictors of future academic and career choices, interest alone does not guarantee that once people are in an occupation, they will find it satisfying.

Strong Interest Inventory. One of the most widely used interest inventories is the *Strong Interest Inventory*® (SII®; Donnay, Morris, Schaubhut, & Thompson, 2005), which can trace its roots to the 1927 publication of the Strong Vocational Interest Blank®. The Strong

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Interest Inventory® assessment, today often simply called the Strong, was called the Strong-Campbell Interest Inventory before 1985. The evolution of this instrument over the past 80 years has resulted in a widely used and respected instrument (Jenkins, 2009). Not only is the Strong Interest Inventory® commonly used in career counseling (Watkins, Campbell, & Nieberding, 1994), but it is also often cited as one of the most widely used instruments in counseling in general (Bubenzer, Zimpfer, & Mahrle, 1990; Elmore, Ekstrom, Diamond, & Whittaker, 1993; Petersen, Lomas, Neukrug, & Bonner, 2014). In addition, the Strong inventory is one of the more researched instruments in counseling, with hundreds of studies having been performed.

The Strong Interest Inventory® compares individuals' responses to items with the response patterns of people in different occupations. This tool is appropriate for high school students, college students, and adults. The 2005 version of the Strong Interest Inventory® contains 291 items, which are at the eighth- to ninth-grade reading level. Individuals respond to items using a 5-point scale (i.e., Strongly Like, Like, Indifferent, Dislike, Strongly Dislike). The items on the 2005 edition are in six sections that address preferences in occupations, subject areas, activities, leisure activities, people, and about their characteristics. Jenkins (2009) recommended that clients be instructed regarding the difference between liking something and being good at it before they take the SII as the assessment is designed to be a measure of interests.

Users can choose from a number of different Strong Interest Inventory® reports such as ones geared toward college or high school students. Before interpreting a Strong report, a counselor should consider the validity of the profile. This can be accomplished, to a degree, by checking the Item Response Percentages. The first step is for the clinician to check the client's total number of responses. If the client answers less than 276, a report will not be generated for that client (Donnay et al., 2005). The second check of a profile's validity would be to examine the Typicality Index, which measures the number of unusual responses given by the client. There are 24 paired items that one would expect individuals to respond to similarly (e.g., liking the occupation of actor/actress and also liking the activity of acting). If the Typicality Index is less than 17, then the counselor should explore why there are so many unusual responses. If an explanation cannot be identified, then the counselor should be cautious in interpreting the results. The third check on the profile's validity involves analyzing the Item Response Percentages. The Item Response Percentages are examined to determine if there are trends such as high percentages of Dislike and Strongly Dislike responses or high percentages of items that were marked Indifferent. Extreme response percentages can be useful in interpreting the results and in providing clinical information.

The information on the Strong profile is designed so that interpretation can progress from general to more specific information. Donnay et al. (2005) suggested explaining the Strong's purpose to the client before interpreting the results so that the client understands that the instrument explores interest not abilities. In discussing interpretation of the 1994 version of the SII, Prince and Heiser (2000) encouraged counselors to provide an overview of Holland's (1997) theory early in the interpretation process. The interpretation should focus on three areas: the General Occupational Themes (GOTs), the Basic Interest Scales (BISs), and the Occupational Scales (OSs). Donnay et al. (2005) stated that after discussing Holland's theory, the counselor should interpret the client's results on GOTs and their results in the six interest areas of Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. After discussing the GOTs, the counselor should then relate the scores and meanings of the BISs under each theme. The 30 BISs can be viewed as subdivisions of the GOTs. For example, under the Social theme—which is related to helping, instructing, and caregiving—the BISs are Social Sciences, Counseling and Helping, Religion and Spirituality, Human Resources and Training, Healthcare Services, and Teaching and Education.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The BISs can help clients understand their underlying interests as related to the six GOTs. For both the GOTs and BISs, the *standard scores* are determined using a norming sample combined of both men and women. However, the distinctions Very Little, Little, Moderate, High, and Very High are determined using either male or female samples. As will be discussed later, there are differing opinions on addressing nontraditional interests for men and women and whether to use combined or separate norms.

In the spirit of going from general to specific, after interpreting the more general interest measures, the counselor can begin to explore the specific OSs. These scales reflect the similarity of the client's interest to the interests of individuals actually working in different occupations. Using people who actually work in the occupations is one of the strengths of the Strong. The current OSs in the SII compare interests of the client to either men or women working in 122 occupations. One of the goals of revising the Strong Interest Inventory was to encourage individuals to explore a wide range of occupations, including ones that may have been dominated by one gender; therefore, each occupation has norms for both males and females. In the past, the Strong had been criticized for focusing on professional occupations, and many nonprofessional occupations have been added to the recent version. Occupations where the client's interests are similar have standard scores of 40 or more and these are the occupations that are often explored in counseling. It is, however, often beneficial for the counselor to also identify occupations for which there is very low interest. If there are no occupations of 40 and above, then the counselor should focus on the occupations where there is the highest interest. The SII is designed to assist counselors in facilitating the identification of interest patterns and examining where the client has interests similar to those working in various occupations. The identification of various occupations is consistent with the current labor and economic market where individuals often change occupations many times in their lives.

After the counselor has explored occupations that correspond to the client's interest, they would then examine the Personal Style Scales of the Strong Interest Inventory®. These scales are designed to help individuals explore how they approach working, learning, leading, making decisions, and participating on teams. The names of the Personal Styles Scales are Work Style, Learning Environment, Leadership Style, Risk Taking, and Team Orientation. The Work Style scale indicates preferences related to working with people as compared with working with ideas, data, or things. The Learning Environment scale indicates an individual's preferred learning environment on the continuum from a preference for academic learning to a preference for a more practically oriented, hands-on environment. The third scale, the Leadership Style, reflects whether people like to lead by example or by directing and persuading others. The Risk Taking scale indicates whether the client prefers to "play it safe" and make careful decisions or takes risks and makes quick decisions. The last scale is Team Orientation, which reflects an individual's preferences related to accomplishing tasks independently or working on a team using a collaborative style. These scales should probably be interpreted more cautiously than other sections of the SII because they are comparatively new and have not been researched as thoroughly as the GOTs, the BISs, and the OSs.

Scoring, Norms, and Psychometric Information. The scoring of the Strong Interest Inventory® is somewhat complex. In scoring the profile, the client's responses are compared to various groups. For both the GOTs and the BISs, an individual's responses are compared with a combined General Representative Sample of 2,250 employed adults. The items that measure the 6 GOTs and the 30 BISs were determined by their content and factor analysis. The sample for the 2005 version consists of an equal number of males and females. Furthermore, 30% of the sample is nonwhite and the sample reflects the distribution of racial and ethnic groups in the United States. Nevertheless, Kantamneni and Scheel (2010)

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

criticized the General Representative sample as having low numbers of Native Americans, Latino(a)s, Asian Americans or Asians, and Hawaiian Natives or Pacific Islanders. In terms of reliability, the test-retest coefficients ranged from .80 to .91. Donnay et al. (2005) provided substantial validation information regarding the six themes of the GOT (i.e., Realistic, Investigative, Artistic, Social, Enterprising, and Conventional). The coefficient alphas for the BIS ranged from .80 to .91 and the test-retest coefficients were .78 to .89. The manual (Donnay et al., 2005) presented some validation evidence regarding the BIS scores such as relationship with college major and previous scores on earlier versions of the SII.

The OS were developed using a different norming procedure and a different method of item selection. Items were selected not because of their similarity in content to other items but rather because of their ability to differentiate the interests of those in an occupation from the general sample (i.e., the General Representative Sample). Thus, the scoring of the OSs is built on comparing the interests of people in certain occupations to a broadly based group of the same gender. Items that differentiate between the occupation and the General Representative Sample were identified, and then weights were assigned based on further analyses. The 2005 version has an adequate sample so that interests can be differentiated for men and women separately for all 122 occupations. Sometimes this required oversampling to ensure an equal number of women and men in each occupation and to include as many individuals from various backgrounds as possible. Kelly (2010), however, noted that some of the data gathered on the OSs was collected 25 or more years ago.

To be included in the norming sample for *each* occupation, an individual had to (1) be satisfied with his or her work, (2) have at least three years of work experience in the occupation, and (3) perform the job duties that are typical and that clearly represent the occupation. The reliability coefficients for the OSs varied with the range of times between testings being from 2 to 23 months and the reliability coefficients ranged from .71 to .91. In terms of concurrent validity, Donnay et al. (2005) found that the male and female scores for many of the occupations were highly correlated and they examined overlapping scores from different occupations. In terms of predicting occupational direction, the 2005 version of the Strong Interest Inventory is different from previous editions, and readers are encouraged to continue to monitor the accumulation of validation evidence on the OSs.

The Personal Style Scales have many features similar to the GOTs and the BISs; however, these scales were first published in the 1994 version of the SII. The Team Orientation scale seems to have the lowest reliability coefficients (e.g., .70 and .74). The Work Style scale has coefficients that range from .90 to .91.

The Strong can be taken online or clients' responses can be mailed in. The Strong Interest Inventory® is often considered a model for other interest inventories in terms of its psychometric characteristics. Certainly, it is the most researched of the interest inventories, and its long history has resulted in a well-developed inventory. There are, however, certain limitations to the Strong, and counselors need to consider whether it is appropriate on a case-by-case basis. For example, Kantamneni and Scheel (2010) question its appropriateness with international samples.

Career Assessment Inventory. In the past, the *Career Assessment Inventory* (CAI) was often considered the "working person's Strong." Whereas previous editions of the Strong Interest Inventory® focused primarily on professional occupations, CAI—Vocational Version (CAI-VV), which was the original version of the inventory, focused on careers that require two or fewer years of postsecondary training. The newer version, the CAI—Enhanced Version (CAI-EV), focuses on careers requiring various levels of postsecondary education. The profile of the CAI-EV is patterned after the Strong. Clients receive scores on General Themes (Holland's six personality types), Basic Interests, and OSs. There are also

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Administrative indexes and Nonoccupational scales (similar to the Strong's Personal Style Scales).

Although in many ways similar to the Strong Interest Inventory®, the CAI has considerably smaller norming samples (e.g., Machinists has 70 males and no females). Miner and Sellers (2002) criticized the use of combined gender norms in the occupational scales of the CAI. The difficulties associated with combined gender norms are discussed in a later section of this chapter. Furthermore, Miner and Sellers suggested that the CAI needs to be updated, in terms of the items and the occupations covered. They also noted a significant shortage of predictive validity studies.

Self-Directed Search. The *Self-Directed Search, Fifth Edition* (SDS; Holland & Messer, 2013) is another inventory that is usually designated as an interest inventory, although its focus is not exclusively on interests. The SDS also includes items related to individuals' abilities and competencies. The Self-Directed Search is somewhat different from other interest inventories as it is designed to be self-administered, self-scored, and self-interpreted. The conceptual base of the SDS is Holland's theory in that the instrument measures the six basic personality types he proposed. The assessment book is accompanied by an *Occupational Finder* that facilitates the exploration of occupations related to the three-letter summary code produced by taking the SDS. The Occupational Finder contains 1,300 occupations that have been linked to the SDS three-letter summary codes. There is also online version and computerized reports. Ciechalski (2009) indicated the SDS is a well-developed instrument with numerous studies to support its use.

Unlike most instruments, the SDS uses raw scores in determining results (i.e., the individual counts the responses related to the six personality types, and the three areas with the most responses produce the three-letter summary code). Daniels (1994) criticized the use of raw data because the four sections of the instruments (i.e., Activities, Competencies, Occupations, and Self-Estimates) vary in the number of items that contribute to the raw scores. In addition, there does not seem to be a theoretical or psychometric reason for the disproportionate contributions to the three-letter summary codes. Although percentiles can be calculated for the SDS, this is not a part of the self-scoring instructions.

In conclusion, the SDS can be an inexpensive and relatively easy assessment device to use. Manuele-Adkins (1989) recommended that when using the instrument, counselors be more involved than the manual recommends. Holland and Messer (2013) suggest that the SDS should only be given to people who want to explore and understand their occupational possibilities and that counselors should use other measures with clients who are less motivated. There are also differences between how occupations are coded between the Occupational Finder of the SDS and those on the Strong Interest Inventory®, which can be confusing for clients who may receive results from both instruments. The SDS scoring is also criticized because it does not provide any consideration of gender differences, so nontraditional interests may be underrepresented. Finally, in my opinion, the validation information on the SDS is supportive of using the instrument for career exploration activities, but other assessment tools and information should be incorporated into the counseling process if the goal is career selection or decision making.

Other Interest Inventories. Counselors use many other interest inventories in addition to the ones mentioned previously. Within the career exploration materials compiled by the U.S. Department of Labor, there is an interest assessment that is related to other O*NET products. The O*NET Interest Profiler can be taken at http://www.onetcenter.org/CIP.html?p=3. The O*NET Interest Profiler measures interests by using Holland typology (i.e., Realistic, Investigative, Artistic, Social, Enterprising, and Conventional). The assessment is designed

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

to be self-administered and interpreted, although current research has found that clients gain more when counseling interventions are utilized with counselor-free vocational assessment (Whiston, Brecheisen, & Stephens, 2003). The website also contains information for counselors on how to effectively use the O*NET Interest Profiler. Pope (2009) recommended the use of the O*NET Interest Profiler with junior high students, high school students, and adults because of its psychometric qualities and because it is inexpensive or free.

Another instrument is the *Jackson Vocational Interest Survey* (JVIS; Jackson, 1996), which is intended for use with adolescents and adults. The JVIS is unique in that it measures preferences in both work styles and work roles. Work styles concern preferences for work environments and include scales such as Dominant Leadership, Job Security, and Stamina. Roszkowski (2001) recommended that the instrument be used with highly motivated clients because of the length of the survey.

Very few interest assessments are published for elementary students. Examples of inventories geared for these younger students are the *Career Finder*, the *Judgment of Occupational Behavior-Orientation* (JOB-O), and the *What I Like to Do* (WILD).

Another method of identifying and assessing career interests is through the use of card sort techniques. In addition to the many commercially available card sorts, practitioners have also developed their own (Hartung, 1999). Hartung concluded that many of the established card sorts are, to varying degrees, theoretically tied to Holland's theory and to the RIASEC (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional) model. Most often, card sorts have occupations listed on cards. The client sorts the cards into three piles: Would Choose, Would Not Choose, and No Opinion. As Williams (1978) contended, card sorts are a creative way of extracting meaningful information from clients, with the added advantage of providing immediate feedback. Card sorts allow for an interactive process between the client and counselor, enabling the counselor to inquire into the reasons for sorting the cards into specific piles. Slaney and MacKinnon-Slaney (1990) maintained that the research on vocational card sorts suggests that they are as effective as other career assessments. The number of research studies on which this conclusion is based, however, is quite small. It may be better to consider card sorts as a supplement to other career assessments. The research does indicate that card sorts should not be self-administered and that the interaction between the client and counselor contributes to the usefulness of the technique.

Some counselors work with clients who have disabilities and special needs. For these clients, the more traditional interest inventories may not always be appropriate. The *Wide Range Interest-Opinion Test-Second Edition* (WRIOT-2; Glutting & Wilkinson, 2003) is a pictorial inventory. Out of a series of three pictures, the clients select the one they like most and the one they like least. Because the instrument does not involve reading, it may be appropriate for clients for whom reading may be a problem. It should, however, be used cautiously with special population as it was normed with a sample from the general population. Other instruments geared for special populations are, the *Ashland Interest Assessment*, the *Geist Picture Interest Inventory*, and the *Reading-Free Vocational Interest Inventory, Second Edition*.

Conclusion. According to MacCluskie, Welfel, and Toman (2002), there can be an artistic quality in explaining the results of an interest inventory. For interpreting the Strong Interest Inventory®, Prince and Heiser (2000) suggested beginning with an explanation of Holland's theory. Savickas, Taber, and Spokane (2002) found that interest scales intended to measure Holland's typology shared only about 35% common variance; therefore, counselors should not assume that the definitions of the scales are the same. Hence, although counselors should not rely on Holland's definitions, they do need to understand how the scales are defined within the specific interest assessments they are using. When interpreting any

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

interest inventory, counselors also need to be vigilant about ensuring that the client understands that the results represent what the client is interested in but do not represent what the client may be good at. Sometimes, particularly when clients are viewing an attractive computer-generated profile, they believe the profile reflects what their future career direction should be. Counselors need to reinforce that the results of an interest inventory are just one piece of information to be used in the career decision-making process.

LO2 Abilities/Skills

In career counseling, Ryan Krane and Tirre (2005) asserted that the assessment of abilities and skills often is conducted to identify occupational possibilities in which the client could be successful. Many of the instruments discussed in Chapter 10 that relate to the assessment of aptitude and achievement can be used in career counseling. Certainly, the Armed Services Vocational Aptitude Battery (ASVAB) and WorkKeys have been used for career exploration and decision making. In addition, these instruments have the option of being used in combination with an interest inventory. Therefore, in considering career assessments, counselors should include those measures mentioned in the previous chapter. Aptitude tests are often used in career counseling because they are good predictors of occupational success (Hunter & Schmidt, 1983). This does not mean, however, that aptitude test results should be examined in isolation as the sole determinant of a client's potential in an occupation. Counselors need to gather information about the whole person when assisting clients in their career planning process.

When interpreting aptitude assessment in career counseling, it is important for counselors to verify the results with other information. Verifying the results entails asking the clients if scores are consistent with other information, such as grades, work performance, self-perceptions of abilities, and other types of evaluations. This verification of the results accomplishes two goals. First, it provides a method of identifying any problems that may have occurred during the testing (e.g., the client's ability to read may have influenced performance, or the client was ill when taking the instrument and the results don't reflect his or her ability). Second, the verification process brings more client information into the interpretative process. Thus, the verification process can assist in providing a more holistic view of the client's abilities as well as giving information on other facets of the client's life. Some aptitude publishers have responded to counselors' need to gather more complete client information by combining an aptitude test with an interest inventory. An example of this is the *Unisex Addition of the ACT Interest Inventory* (UNIACT), which is a key component of the ACT. The combination of aptitudes and interests facilitates exploration and leads to a more comprehensive view of the client.

Campbell Interest and Skills Survey. A different method of combining interests and skills or ability assessment is available through the *Campbell Interest and Skills* (CISS; Campbell, 1992). Campbell, the author of this instrument, has a long, distinguished career in vocational assessment and was instrumental in the development of the Strong Interest Inventory's® predecessor, the Strong-Campbell Interest Inventory. The CISS combines the self-assessment of skills with an exploration of interest. This combination encourages the client to examine the interaction between interests and skills and to examine career areas in which the two intersect. As is discussed later, research supports the use of self-estimates of skill level assessment.

The 320 items of the CISS result in scores on 7 Orientation Scales, 29 Basic Interest and Skills Scales, and 58 Occupational Scales. The Orientation Scales roughly correspond to Holland's six themes, except that the Realistic theme is broken down into two categories, the Producing and the Adventuring orientations. The Basic Scales are subcategories of the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Orientation Scales and represent various occupational activities. Unlike some of the other interest inventories, the norms for these two scales and the Occupational Scales are unisex rather than reporting scores for men and women separately. The CISS also includes three procedural checks (a Response Percentage Check, an Inconsistency Check, and an Omitted Items report).

Based on the scores on interests and self-assessment of skill level, the profile provides four categories: Pursue—high interest and high skills; Develop—high interest and lower skills; Explore—lower interest and high skills; and Avoid—lower interest and lower skills. Pugh (1998) recommended that counselors use caution in interpreting the Develop and Explore designations because a small difference in interests and skills can result in different categorizations. The reference sample of 5,225 seems adequate; however, the sample has substantially more men than women. The median coefficient alpha is .87 for the Orientation Scales and .86 for the Basic Scales. There is more variability in the OSs, and the skills assessments on the OSs tend to be less reliable than the interest scores. Validation evidence is growing; it is anticipated that in 10 years, there will be substantially more. Hansen and Leuty (2007) found convergent evidence related to the skills portion of the CISS and that the scores were predictive of different college majors. The interest assessment has a strong empirical base because it comes from Campbell's work on the Strong-Campbell Interest Inventory.

Skills Confidence Inventory. With the revision of the Strong Interest Inventory®, there has also been a revision of the *Skills Confidence Inventory* (SCI; Betz, Borgen, & Harmon, 2005). The SCI measures perceptions of one's capabilities in performing activities that correspond to Holland's six interest themes and it is typically used in conjunction with the Strong Interest Inventory®. This 60-item instrument is designed to measure an individual's self-perceived ability or self-efficacy expectation in successfully completing a variety of tasks, activities, and coursework. Scores on the six scales range from 1 to 5. The SCI was designed to be used in conjunction with the Strong Interest Inventory® so that individuals can examine their interests and perceived abilities simultaneously.

Self-Estimates of Abilities. In considering methods for assessing skills and abilities, counselors may want to consider the use of self-estimates. Westbrook, Sanford, Gilleland, Fleenor, and Merwin (1988) found a significant relationship between self-appraisal estimates of abilities and measured abilities. They also found, however, that adolescents generally tended to overestimate their abilities. Their research indicates that counselors can use self-appraisals in career counseling. Prediger (1999) recommended using "informed self-estimates" of abilities, which draw on clients' experiences.

In the literature on self-estimates of abilities, there tends to be two constructs: ability self-estimates and self-efficacy estimates (Bubany & Hanson, 2010). As an example, the CISS uses self-estimates of abilities, whereas the SCI uses self-efficacy estimates. In considering self-estimates and self-efficacy, Tracey and Hopkins (2001) suggested an important distinction between them. They contended that self-efficacy beliefs concern future performance with the client possibly using a criterion-referenced framework. On the other hand, self-estimates tend to involve current levels of abilities and often involve an implicit assumption that one is comparing one's abilities to others. The research regarding whether they are distinct is mixed. Brown, Lent, and Gore (2000) found self-estimates of abilities and self-efficacy estimates to be correlated, but they concluded they were distinct constructs. In contrast, Bubany and Hanson (2010) found that any difference between the two constructs was due to measurement error or scale content. In conclusion, although there may be some distinctions between self-estimates and self-efficacy, there is also probably some overlap. In working with either self-estimates of abilities or self-efficacy estimates,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the counselor should validate the results by examining whether they are consistent with other information about the client (e.g., grades, other assessments).

LO3 Values

In the area of career assessment, the exploration of values is sometimes neglected, even though it is often a critical aspect of the career counseling process. Rounds and Armstrong (2005) contended that values should play a vital role in career decision making. Osborne, Brown, Niles, and Miner (1997) indicated that "The understanding of an individual's values is, perhaps, the most important element of the decision-making process. Values serve as the basis for which life goals are established and provide a guide for individuals in making personally consistent and meaningful decisions about the future" (p. 82). Work values are more highly correlated than interest with work satisfaction (Rounds, 1990). Some of the instruments typically associated with the assessment of values in the career area use terms such as *needs* or *preferences.*

If a counselor is using a values inventory, the practitioner should remember that no inventory is inclusive of all possible values. In other words, a client may value something that is not assessed on the instrument being used. Chartrand and Walsh (2001) argued that those career assessments in which the focus is only on work values are insufficient because many people may want to examine their values across multiple life roles. Counselors need to supplement the use of a values inventory with an exploration of other possible values, which can be accomplished by asking the client questions about personal values not otherwise assessed or by exploring the client's lifestyle.

Minnesota Importance Questionnaire. The *Minnesota Importance Questionnaire* (MIQ; Weiss, Dawis, & Lofquist, 1981) relates to the theory of work adjustment but is also used as an assessment of values or needs in career counseling. It measures 20 vocational needs (e.g., Ability Utilization, Compensation, Coworkers, Recognition) and six underlying work values (Achievement, Altruism, Autonomy, Comfort, Safety, and Status). Needs are indicated by the number of times an individual selects a statement as being important relative to other statements, and values are defined as a cluster of related needs. Brooke and Ciechalski (1994) rated its psychometric qualities positively and the instrument has been heavily researched.

Values Scale. The *Values Scale* (Nevill & Super, 1989) is a measure of both work-related values and general values. The 106 items produce scores on 21 values (e.g., Ability Utilization, Achievement, Economic Security, and Personal Development). The authors of this scale have labeled it the Research Edition to indicate that clinicians should be cautious in using it. There is a normative sample, but the authors recommend using an ipsative interpretation. With *ipsative interpretation,* the variables are compared with each other without reference to whether the client's values in that area are high or low as compared with values of other people. Slaney and Suddarth (1994) wondered if this cautious approach is even warranted given the lack of predictive validation studies; however, Schoenrade (2002) considered the content validation evidence to be respectable. Hackett and Watkins (1995) agreed that there were drawbacks with the instrument but suggested it had cross-cultural applications and did not appear to be gender-biased. The Value Scale is one of the instruments administered in Super's Career Development, Assessment, and Counseling (C-DAC) model (Osborne et al., 1997).

According to Rounds and Armstrong (2005), although assessment of values will often provide clients with important information to consider in career planning, values measures have not received the same attention as other typical career counseling assessments

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(e.g., interests inventories and ability assessments). They recommended that values assessment be used in a comprehensive strategy that involves occupational information.

Integrative Career Assessment Programs

As noted earlier, a number of interest and aptitude assessments are now offered in conjunction with each other. This section discusses a few of the programs that combine interests, abilities, and values assessments.

Kuder Career Planning System. Building on the historical work of Fredrick Kuder, the Kuder® Career Planning System (KCPS) is an online, interactive program that is the latest in a series of career assessment tools that dates back to the 1930s. The KCPS includes the *Kuder Galaxy* for elementary students, the *Kuder Navigator* for middle school and high school students, and the *Kuder Journey* for postsecondary students and adults. Kuder Galaxy is not a series of career assessments; it is a set of interactive games and activities to help students at each grade level between pre-kindergarten and fifth grade learn about themselves and the world or work. With Kuder Navigator and Kuder Journey, individuals can take the Kuder® Career Interests Assessment™, the Kuder® Skills Confidence Assessment™, and the Kuder® Work Values Assessment™. These three assessments are only part of the system for each age level. Other activities such as exploration of occupations, goal setting activities, and developing electronic portfolio are parts of these comprehensive career guidance programs as well. According to Schenck (2009), one of the greatest strengths of the Kuder programs is the extensive information infused into the Kuder website. These programs can be purchased online by individuals or schools, community agencies, and other organizations.

The interest inventory, the *Kuder® Career Interests Assessment™* (KCIA), is a 32-item assessment that measures Holland's six themes (Kuder, Inc., 2012a). From Holland's six themes, the individual is then directed to O*NET occupations. When taking the KCIA, individuals select from three activities that they most prefer, that they second most prefer, and that they third most prefer. In the system for adolescents (i.e., the Kuder Navigator), respondents are also referred to clusters from the 16 national career clusters of pathways. School counseling students should become familiar with the 16 national career clusters or pathways as they are often used in career advising in the schools. The reliability coefficients for the six Holland themes ranged from .73 to .87. Researchers also analyzed each item for group sensitivity and either removed or modified any item that varied by 20% or more.

The *Kuder® Skills Confidence Assessment™* (KSCA) is for Kuder Navigator, whereas the *Kuder® Skills Confidence Assessment-Adult™* (KSCA-A) is for Kuder Journey. Both of these assessments measure relative self-efficacy related to Holland's six themes. Similar to the *Kuder® Career Interests Assessment™*, these self-efficacy scores are used to identify O*NET occupations and clusters from the 16 national career clusters or pathways. For the KSCA, the reliability coefficient values ranged from .74 to .91. The reliability coefficient were a little higher for the KSCA-A and ranged from .85 to .93. For more information on the psychometric qualities of the KSCA and the KSCA-A, see Kuder, Inc. (2012b and c).

The *Kuder® Work Values Assessment™* (KWVA) measures five work-related values: *Workplace*, *Innovation*, *Accomplishment*, *Income*, and *Prestige* (Kuder, Inc. 2012d). This instrument is based originally on Super's Work Values Inventory-Revised. Reliability coefficients for the five values ranged from .83 to .89. Furthermore, other psychometric and validation evaluations involved a sample of 4,000. Gibbons (2013) evaluated all three of the assessments positively and indicated that they were a sound and valid way to help clients explore career-related information. One of the advantages of the Kuder Career Planning System is that it provides an easy "one-stop" location for assessment related to career development for clients from a variety of ages.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

COPSystem. *The COPSystem* is published by EdITS/Educational and Industrial Testing Services and combines separate interests, abilities, and values inventories. It is one of the few comprehensive assessment programs that is not Internet based but rather a paper-and-pencil assessment. The instruments in this system can be used alone or in tandem. Using the Comprehensive Career Guide, the results from the interests, abilities, and values assessments can be interpreted, with the results from all three assessments using the same occupational classification system. Depending on the client's needs, the practitioner can select from several versions of the COPSystem Interest Inventory (COPS), such as the general form (COPS-Revised), a professional focused instrument (COPS-P), an instrument appropriate for elementary through high school (COPS-II), and a pictorial instrument (COPS-PIC). The values instrument, the Career Orientation Placement and Evaluation Survey (COPES), was revised in 1995. The COPES instrument measures eight values using a continuum (e.g., Investigative vs. Accepting, Practical vs. Carefree, Independence vs. Conformity). Thus, counselors can use the results to interpret which area is valued over the other on the continuum and identify from the eight areas the ones that are most valued. The ability measure is the Career Ability Placement Survey (CAPS), which consists of eight subtests. Each of the subtests lasts only five minutes and, therefore, is not a comprehensive test of those eight abilities. Each instrument in the COPSystem can be hand scored by the client, and there are procedures for integrating the results of the three instruments to facilitate a more comprehensive exploration experience. It can also be machine scored; counselors can purchase the Integrated Report Interpretation System (IRIS), which provides computer-generated reports. Although the COPSystem continues to be revised and improved, there is limited validation evidence, particularly longitudinal studies. The COPSystem can provide an integrative program for career exploration, but sometimes clients will benefit from more thorough assessments. Although the COPSystem is designed as a self-directed exercise in which clients can take and interpret their results unaided, Wickwire (2002) strongly encouraged counselors to be actively involved in the process.

Combining Assessment Information

Integrative and comprehensive career assessment does not have to involve only one system, such as the COPSystem; counselors can integrate a number of separate career assessments. The Career Development Assessment and Counseling Approach (C-DAC; Osborne et al., 1997) involves integrating the results from a number of career assessments: the Adult Career Concerns Inventory, the Career Development Inventory, the Strong Interest Inventory®, the Values Scale, and the Salience Inventory. In addition, career assessment does not always have to involve formal assessment instruments; some informal exercises have been shown to be valid tools (e.g., the Occupational Daydream exercise in the Self-Directed Search). Integrating information on interests, abilities, and values is a common part of the career counseling process. Armstong, Day, McVay, and Rounds (2008) stressed the importance of integrating individual career assessment using an organizational framework. They proposed the Atlas Model of Individual Differences that starts with interest and uses Holland's (1997) typology as the framework for interpreting personality, skills, and values. In terms of integrative career assessment, Spokane and Jacob (1996) suggested a portfolio approach that could facilitate the integration of interests, abilities, and values assessment. Portfolio assessment also makes sense in terms of career development because students can keep career counseling exercises and results of assessments in their portfolios. School counselors could then use the portfolios to integrate career information gathered from students throughout their academic careers. Hence, an interest inventory taken in middle school could be compared with an interest inventory taken in high school.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

LO4 Career Process Measures

The previous discussion addressed assessments of individual differences that are used to assist clients in career planning or addressing various issues related to work. These instruments assess various content areas (e.g., interests, values) that have been shown to be related to sound career choices. This chapter's focus is now moving toward instruments that measure where clients are in the career development or decision-making process. Typical measures of career process variables are related to level of decidedness or certainty and an individual's level of career maturity/adaptability.

Career Decision Making

It is often helpful in career counseling to have an indication of where the client is in terms of career decision or indecision. The interventions a counselor selects would probably be very different for a client who has not made a career decision as compared with individuals who are decided. A distinction should be made between clients who experience normal developmental undecidedness and those who experience indecisiveness. The term *undecided* means the client has not yet made a career decision, whereas the term *indecisiveness* indicates chronic problems with indecision (Betz, 1992). A number of instruments have been found to be useful in assessing clients' progress in the decision-making process or in identifying difficulties clients may be experiencing in making a career decision.

A widely used instrument in this area is the *Career Decision Scale* (CDS; Osipow, 1987). The CDS includes 19 items. The first two items measure the level of certainty the individual feels in making a career decision. The next 16 items measure the level of career indecision. The last item provides an opportunity for individuals to describe their career indecision if none of the other items seemed appropriate. The CDS is generally regarded as a reliable instrument, with validity evidence related to its use as a measure of overall career indecision (Harmon, 1994). It is important to note that the CDS provides a measure of career indecision, but it does not indicate the source or the type of indecision. Another measure often used in this area is Holland, Daiger, and Power's (1980) *My Vocational Situation* (MVS). The MVS has three subscales: the Vocational Identity scale, the Occupational Information scale, and the Barriers scale. The Vocational Identity scale is often considered a measure of indecision and assesses the clarity and stability of one's career-related attributes. Both the CDS and the MVS are used not only to gather information about clients but also to evaluate the effectiveness of career counseling.

An instrument that is increasingly being used in career counseling and outcome research is the *Career Decision-Making Self-Efficacy Scale* (CDMSE; Betz & Taylor, 1994). Often used with college students, the CDMSE measures the confidence individuals have in their ability to make career decisions. This instrument draws from Bandura's (1977) theory of self-efficacy, which proposed that an individual's belief about his or her ability to perform a specific task or behavior significantly affects the individual's choices, performance, and persistence in that task. There is also the *Career Decision-Making Self-Efficacy Scale-Short Form* (CDMSE-SF, Betz, Klein, & Taylor, 1996), which consists of only 25 items. There are five scales (i.e., Accurate Self-Appraisal, Gathering Occupational Information, Goal Selection, Making Plans for the Future, and Problem-Solving) in both the original and the shortened versions of the CDMSE. The CDMSE-SF can be particularly useful in evaluating whether career interventions help individuals become more efficacious in their career decision making.

An instrument that can be used in identifying or diagnosing decision-making difficulties is the *Career Decision-Making Difficulties Questionnaire* (CDDQ; Gati, Krausz, & Osipow, 1996). The CDDQ was constructed to measure a theory-based taxonomy of decision-making difficulties. The taxonomy includes three major areas of difficulties that

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

are further subdivided into 10 specific difficulty categories, and, finally, into 44 specific difficulties. For example, the first of the three major dimensions is Lack of Readiness, which is subdivided into the categories of lack of readiness due to Lack of Motivation, Indecisiveness, and Dysfunctional Beliefs. The second major dimension is Lack of Information, with the subcategories being Lack of Information about the Career Decision-Making Process, Lack of Information about Self, Lack of Information about Occupations, or Lack of Information about Ways of Obtaining Information. The last dimension is Inconsistent Information, and the categories are Unreliable Information, Internal Conflicts, and External Conflicts. Studies related to the psychometric qualities of the CDDQ have been mostly favorable (Gati, Osipow, Krausz, & Saka, 2000; Osipow & Gati, 1998).

Career Maturity

Betz (1988) defined *career maturity* as the "extent to which the individual has mastered the vocational tasks, including both knowledge and attitudinal components, appropriate to his or her stage of career development" (p. 80). Career maturity is examined developmentally and compares the individual to others in terms of career development. Essentially, career maturity measures the client's level of readiness for mastering career development tasks. In a discussion of the assessment of career maturity, it is important to note some criticisms of the construct and the assumptions that career development is always temporal, unidirectional, and linear (Savickas, 2001). Career maturity measures, however, are often used to identify limitations or deficiencies in terms of an individual's career maturity and as an evaluation tool to determine if career counseling has increased the client's career maturity.

The *Career Development Inventory* (CDI; Super, Thompson, Lindeman, Jordaan, & Myers, 1988b) measures career maturity using Super's theory. The scales of Career Planning, Career Exploration, Decision Making, and the World of Work correspond to Super's theory, with the exception of not including a measure of Reality Orientation. An overall career maturity measure, called the Career Orientation Total, is also available. The CDI does measure a cogent model of career development, but some have questioned whether the model still has applicability given the rapid changes in the workforce and global labor market. A somewhat related instrument is the *Adult Career Concerns Inventory* (ACCI; Super, Thompson, Lindeman, Jordaan, & Myers, 1988a), which evolved from the Career Development Scale (Adult Form). Super suggested that the term *career adaptability* is more appropriate for adults than the term *career maturity*. The ACCI, which does not follow the same format as the CDI, focuses solely on the planning dimension of Super's model. The ACCI primarily assesses adults' needs for career planning and adaptation, but it should not be used if a counselor desires a measure of adult career competencies, such as decision-making skills or job search skills.

Career Adaptability

Although Super contended that career adaptability was a more appropriate term for adults than career maturity, Savickas (2005, 2011a) argued that in our ever-changing labor market career adaptability is a more germane construct than career maturity. Savickas and Porfeli (2011) defined career adaptability as "a multidimensional construct that characterizes an individual's psychosocial readiness and resources for coping with current and imminent vocational development tasks, occupational transitions, and work traumas" (p. 357). According to Savickas (2005) in career construction theory, students should approach career choice readiness tasks with *concern* for their futures, a personal sense of *control* over their careers, the *curiosity* to explore possible selves and social opportunities, and the *confidence* to design their career futures and execute plans to make these career decisions happen.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

There was a widely used measure of career maturity called the *Career Maturity Inventory* (Crites, 1973, 1978; Crites & Savickas, 1995), which Savickas and Porfeli (2011) took items from to create the *Career Maturity Inventory-Form C*. This new measure applied career construction theory (Savickas, 2005, 2011a) to form a measure of career adaptability for students in 12th grade and below. The Career Maturity Inventory-Form C has five scales with the first on being Career Choice Readiness, which is a total of the other four scales (i.e., Concern, Control, Curiosity, and Confidence). For the Career Maturity Inventory-Form C, the Concern Scale assesses the extent to which an individual is engaged in the process of making career decisions. The Curiosity Scale measures the degree to which individuals are seeking occupational information and exploring the world of work. The Confidence Scale measures the extent to which individuals have faith in their ability to make realistic occupational choices. The fifth score is for the Consultation Scale, which assesses the degree to which individuals seek assistance in career decision making, seeking information and advice from others (Savikas & Porfeli, 2011). The Career Maturity Inventory-Form C is available through Vocopher.

An exciting project that involved researchers from 13 countries produced the *Career Adapt-Abilities Scale* (CAAS), which measures career adaptability in adolescents and adults. Researchers from Belgium, Brazil, China, France, Island, Italy, Korea, the Netherlands, Portugal, South Africa, Switzerland, Taiwan, and the United States all administer the CAAS and evaluated its reliability and validity. The CAAS contains 24 questions that measure the scales of Career Concern, Career Control, Career Curiosity, and Career Confidence. Each scale has six items. When the data from all countries were combined, the reliability coefficients for the four scales were considered good to excellent (Savickas & Porfeli, 2012). The internal consistency estimates for the CAAS total score were excellent across all countries. The CAAS represents one of the best examples of an instrument developed for international use and one that has applicability for use with individual clients.

Other Career Choice Process Measures

Besides measures of career decision and career maturity or adaptability, there are other measures that can be used as part of the career counseling process. One example is the *Career Thoughts Inventory* (Sampson, Peterson, Lenz, Reardon, & Saunders, 1996), which is designed to measure dysfunctional thinking about career decision making. The Career Thoughts Inventory can assess whether the individual is having difficulty sustaining a career decision, whether there is difficulty in making a commitment to a career, or whether external forces (e.g., parents) are in conflict with self-perceptions. Another possible instrument is the *Career Transitions Inventory* (Heppner, 1991), which is designed to assess clients' perceptions of psychological resources available to individuals as they are going through a career change.

As you may have noticed, many of the assessments that have been used to assess clients' career decision making, career development or maturity, or other career counseling process measures were published in the 1980s and 1990s. Another potential problem concerns whether existing instruments are sensitive to multicultural issues and can be used with clients from diverse backgrounds. Therefore, there is a need for new measures to use with clients in determining the need for certain types of career counseling, and for use in program evaluation.

LO5 Qualitative Career Assessment

Some counselors have questioned whether career assessment attempts to objectively measure clients' traits are consistent with postmodern philosophy and constructivism (McMahon, Patton, & Watson, 2003). One of the first professionals to promote qualitative assessment was Goldman (1990), who contended that qualitative assessment can most

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

easily be defined by what it is not. Goldman contended that qualitative assessment is for the most part not standardized tests that usually yield quantitative scores and norm-based interpretation. In Goldman's view, qualitative assessment tends to foster a more active role for the client rather than a more passive interpretation of the results provided by the clinician. Goldman also suggested that qualitative assessment emphasizes a holistic study of the individual compared with relying on more discrete measures of human constructs such as interest, abilities, or personalities.

Qualitative career assessment is not a set of specific assessment instruments or techniques; however, there are some commonly used techniques. In my opinion, the most commonly used qualitative career assessment is the Career Story Interview by Savickas (2011b). In order to conduct the Career Story Interview, the counselor must understand career construction theory (Savickas, 2011b). The questions asked during the Career Story Interview are (1) How can I be useful to you as you construct your career? (2) Who did you admire when you were growing up? Tell me about him or her. (3) Do you read any magazines or watch any television shows regularly? Which ones? What do you like about these magazines or television shows? (4) What is your favorite book or movie? Tell me the story. (5) Tell me your favorite saying or motto. (6) What are your earliest recollections? I am interested in hearing three stories about things that you recall happening to you when you were 3 to 6 years old, or as early as you can remember. These questions are not interpreted in a vacuum, and career counselor needs to be trained on how to make meaning of the client's answers. Interested readers are directed to Savickas (2011b) who provides a very thorough overview of how to conduct and interpret the Career Story Interview.

Chope (2015) provides some other examples of qualitative career assessments, such as card sorts, sentence completions, career genograms, and lifelines. Card sorts have already been discussed briefly, but card sorts can also be customized to address other topics than having the client sort occupations such as values, challenges and barriers, and other career-related topics. Sentence completion exercises will be discussed in Chapter 12, but these can also be adapted to address career issues. An overview of genograms is provided in Chapter 13, but genograms can also be adapted to focus on career issues. Lifelines are a commonly used technique in which the client demarcates significant times or events in his or her life. Typically the clients draw a straight line on a piece of paper, which represents an average level of functioning. Life and work experiences that the client views as positive, happy, or rewarding are plotted chronologically above the line, whereas experiences that are negative or painful are plotted chronologically below the line. The counselor then works with the client on interpreting these experiences.

Cohen, Duberty, and Mallon (2004) addressed the importance of social and cultural influences and themes of power and ideology in research using career narratives. Many of these same issues related to attending to social contextual factors as they apply to the use or development of qualitative career assessments. These authors suggested that the construction of narratives needs to elucidate the socially and culturally embedded nature of career and facilitate a greater understanding of the relationship between the individual and the social context. Hence, when one is using a qualitative career assessment, an emphasis would be placed on social phenomena and the interrelationship with the cultural and historic context of the individual.

To use qualitative career assessment, a practitioner must understand the philosophical underpinnings of the approach. The philosophical foundations of a narrative approach are different from most career development theories that often have the goal of *predicting* career choice or career-related behaviors (Bujold, 2004). This philosophical distinction is paramount, because in traditional career assessment, a clinician would focus on the degree to which the instrument provides assistance in determining a career direction. With qualitative career assessments, the focus is on postmodern philosophy and the idea

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

of cocreating a career narrative. Nevertheless, Sampson (2009) argued that postmodern and modern approaches are not incompatible and that clinicians can use career assessment from both philosophical approaches.

Issues and Trends in Career Assessment

In the area of career or vocational assessment, it is particularly important for counselors to consider current trends and emerging issues. Often in vocational assessment, counselors are assisting individuals in planning their work lives, and with the rapid changes that are occurring in our society (Debell, 2006) this is becoming increasingly more difficult. Technology continually influences the counseling field and this is particularly true in the area of vocational assessment. Many of the instruments that have been discussed in this chapter can be taken over the Internet and clients can receive results within seconds. This evolution in career assessments has many advantages, but there also are some disadvantages and some issues that counselors should consider related to Internet-based career assessment. Other issues concern gender and racial group differences regarding career exploration and planning. In using career assessments, counselors need to be aware of issues related to gender, race, and ethnicity to best serve the needs of their clients. Although this is true in every area of assessment, it is particularly true in the area of career assessment where some clients are likely to see the results of a career assessment as being prescriptive and not understand the limits of these instruments.

LO6 Technology and Internet-Based Career Assessments

The use of computers has become ingrained in many aspects of vocational assessment (Sampson, Carr, Lumsden, Smission, & Dozier, 2009). The number of sites on the Internet expands exponentially each day, and this same expansion has occurred in websites on career assessment. In examining instruments included in *A Counselor's Guide to Career Assessments Instruments* (Kapes & Whitefield, 2002), Sampson, Lumsden, and Carr (2002) found that only about 18% of career assessments were available directly on the Internet, whereas today many of the commonly used assessments have Internet versions that can be purchased by individual clients or counseling organizations. The ability for individuals to purchase career assessments directly is problematic. As Whiston, Brecheisen, and Stephens (2003) clearly demonstrate, counselor-free career interventions are not particularly effective and significantly less effective than those involving a counselor.

In one of the first attempts to examine career counseling Internet sites, Oliver and Zack (1999) evaluated 24 no-cost career sites. These researchers found that the content of the sites varied markedly, with many sites deficient in terms of the quality of their assessments and interpretations. They found only one site that reported psychometric data for its measures, and information concerning the development of the sites was not attainable, even when efforts were made to contact the site developers. Oliver and Zack concluded that although they saw considerable potential in the Internet sites, the sites had many shortcomings. My perusal of many online-based career assessments is that many of them have enhanced graphics but they do not provide psychometric information on the assessments.

Barak (2003) contended that the area of online career assessments may be divided into two groups. The first group includes assessments provided by professionals intended for use by individuals or groups, such as the Self-Directed Search online version (http://self-directed-search.com). The second group includes amateur tests, quizzes, questionnaires, and illegally adapted copies of existing instruments. Even for existing instruments that were developed by professionals, many authors emphasize the necessity of validating career counseling instruments and interpretive materials now found on websites

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(e.g., Sampson & Osborne, 2015). Although there may be extensive research related to a paper-and-pencil version of an instrument, there still needs to be additional research related to whether the Internet version is comparable to the paper-and-pencil version. For many users of Internet assessments, it is difficult to ascertain whether the instrument is sound and methodologically strong or if it is some quickly developed test with no basis. Although this same criticism can be made of paper-and-pencil assessments, these instruments are typically given by professionals who have the necessary training to use them. The Internet is not regulated; thus, in many ways, the consumer may be purchasing worthless information. In addition to the monetary costs, there are other costs of using online material. For example, an individual might quit a job, enter an expensive training program, or seek unsuitable work based on the results of an instrument developed by a nonprofessional. Hence, Internet career assessment has the potential to harm individuals.

One issue particularly salient to Internet use concerns privacy and keeping results confidential. Although some sites are protected by advanced security procedures, this is not true of all sites. Furthermore, in face-to-face interpretations of test results, the counselor has the opportunity to explain assessment results that may be confusing or upsetting to the client. This ability is completely lost if the interpretative material is provided only over the Internet. Although somewhat dated, the National Career Development Association (NCDA) has developed *Guidelines for the Use of the Internet for the Provision of Career Information and Planning Services* (National Career Development Association, 1997).

Strengths and Limitations of Internet or Technology-Assisted Career Systems. Over the last 20 years, computers and technology have changed career assessments. It may no longer be appropriate to use the simplistic term *computer-assisted career assessment systems* because individuals are now accessing information not just from computers but also from devices other than their computer (e.g., cell phone). In addition, with the explosion of the Internet, career assessments are not necessarily downloaded onto a computer as assessments, and interpretations can often be more efficiently managed through a website. Hence, the term I will use is *technology-assisted career assessments.* Technology-assisted career assessment systems have many advantages. (1) The systems can store large amounts of information that can be easily updated, (2) clients can retrieve information that corresponds to their individual needs, (3) clients can progress through the assessments and information at their own pace, and (4) many of these systems guide clients through a systematic decision-making process that can serve as a framework for future career decisions.

Another problem concerns the perception that these technology- or computer-assisted programs provide "the answer" rather than encouraging the client to explore and make informed career decisions. Because of the nature of computers, some clients are likely to see these programs as being more "scientific" and providing a definitive career decision. Individuals may access career assessments through the Internet and not seek assistance related to interpreting the results. As indicated earlier, counselor-free career interventions have not been found to be effective and clients may make poor choices without the assistance of a trained counselor. Computers and other technologies are well suited to repetitive tasks and computational activities (e.g., administering items, scoring scales, and constructing profiles), which can give clinicians more time to integrate findings, interpret results, and assist clients with career-related issues. In addition, some career assessment sites are incorporating multimedia to deliver career information in a more engaging manner (Sampson et al., 2009), which is a strength of these systems.

In my view, one of the major problems with career assessments on the Internet is the difficulty many individuals may have in telling the differences between legitimate websites and ones constructed in a haphazard manner with little if any attention to the quality of the assessments. Barak and English (2002), in their thorough discussion of the pros and

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

cons of psychological testing on the Internet, have stressed the need for examining the psychometric properties of tests. Reliability and validity data are needed for all online instruments. Some instruments may appear to be legitimate, but professionals may not have critiqued those Internet-based career assessments. Whiston and Oliver (2005) encouraged the involvement of counseling professionals in the construction, maintenance, and operation of Internet sites related to career assessment.

LO7 Gender and Career Assessment

In recent years, one of the issues in career assessment has been related to gender differences on some career assessment instruments. Research related to women's career development abounded in the 1980s and 1990s, and as a result, problems with women's career assessment were identified. According to Harmon (1997), "Women who experienced the career assessment process of the 1960s often report that their nontraditional interests were dismissed by their counselors as inappropriate" (p. 467). Although it is obvious that many changes have taken place in this field since the 1960s, career assessment with women still requires specific knowledge about gender differences and current knowledge of assessment research. Interest inventories typically use questions that are reflective of differential experiences. For example, the answers to questions such as "I like to fix car engines" and "I like to take care of children" may be influenced by whether the individual has had opportunities to participate in these activities. Therefore, it is not surprising that interest inventory results show gender differences. In a meta-analysis of interest assessments of people, data, and things, Su, Rounds, and Armstrong (2009) found that men were more interested in things and women were more interested in people. They also explored sex differences using Holland's typology and found that men showed stronger interests in Realistic and Investigative areas, whereas women showed stronger Artistic, Social, and Conventional interests. They also found that men were more interested than women in engineering, science, and mathematics.

Betz (1992) and Hackett and Lonborg (1993) suggested that interest inventories should use same-sex normative scores rather than combined norms. With same-sex norms, females' interests would be compared with those of other females, which would increase the likelihood of nontraditional interests being identified. For example, a female who is interested in a number of Realistic activities may receive a score of moderately high interest in this area when compared with other females, but her score may appear only as average or even below average if half of the norming group is male. Separate norming groups, however, can have a subtle influence on the manner in which clients view their results, and, thus, counselors need to address sexual stereotyping in their interpretations of interest inventories.

To supplement structured interest inventories, Hackett and Lonborg (1994) proposed the use of less-structured assessment methods, such as value clarification exercises. These less-structured assessments can more easily incorporate discussion of gender and cultural issues. Betz (2005) suggested that any type of interest assessment should be guided by exploratory versus confirmatory objectives and that the interpretation should be in line with the "opportunity dominance" versus "socialization dominance" approach. The opportunity dominance approach emphasizes that low interest scores may be a reflection of an individual's restricted learning experiences and opportunities and that new interests can develop if individuals are exposed to new experiences. Hence, low scores in mechanical and building trade occupations may be a reflection of a lack of exposure rather than solely a lack of interest. Betz (2005) therefore proposed that in addition to using same-sex norms or sex-balanced interest inventory scales, counselors have a responsibility to assist women in broadening their experiences, particularly in areas such as the sciences, engineering, and technology.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In assessing career factors with women, it is important to examine both internal and external barriers. Examples of external barriers include gender-role stereotypes, obstacles in educational systems, a lack of role models, and the "null environment" (Betz, 2005). Higher education is often referred to as a null environment because women are neither supported nor encouraged. Thus, they are in an environment that is void of assistance. Although this external environment is not overtly discriminatory, the lack of encouragement often results in failure. Environmental barriers are not the only barriers—the socialization process for many women results in such internal barriers as low self-efficacy, performance anxiety, and lower aspirations and expectancies. In general, when using either formal or informal career assessment strategies, the counselor needs to be sensitive to gender issues and cognizant of how stereotypes may be affecting the assessment process.

LO8 Ethnic and Cultural Differences in Career Assessment

Although issues related to multicultural assessment are discussed in Chapter 6, some issues are particularly relevant to career assessment. Sensitivity to multicultural context in career assessment is important. Fouad and Byars-Winston (2005) found, in a meta-analysis, differences among racial/ethnic groups in perceptions of career-related opportunities and barriers. Therefore, in interpreting the results of a career assessment, a counselor needs to explore both perceived and actual barriers of individuals of color because these barriers may not be readily apparent to the counselor.

Leong and Hartung (2000) analyzed the literature related to multicultural career assessment and discussed it related to the dimensions of *cultural validity* and *cultural specificity.* Cultural validity addresses the concept of whether instruments that have been developed primarily from a Eurocentric perspective can be used with clients from other cultural backgrounds. Cultural specificity, on the other hand, concerns the extent to which variables, such as worldview, cultural identity, communication style, and decision-making preferences, are addressed in the assessment process.

Career researchers have also examined the appropriate use of interest inventories with racial and ethnic minorities. Although early research cast some doubt on the use of Holland's hexagonal model with ethnic minority samples in the United States (Rounds & Tracey, 1996), later research supported a common vocational interest structure among racial and ethnic minorities (Day & Rounds, 1998; Fouad, Harmon, & Borgen, 1997). The Day and Rounds results are particularly noteworthy because their sample was more than 49,000 and involved large groups of minority students. Their multidimensional scaling analysis provided support for traditional methods of assessing career interests by indicating that individuals from diverse ethnic groups use the same cognitive map or structure of preference when examining career interests. This does not mean, however, that the percentage of individuals from different ethnic groups will be the same for Holland's six areas because we do see that certain themes tend to have higher representation. For example, African Americans are more likely to have high Social interests, and Asian Americans tend to have higher Investigative interests.

As indicated in Chapter 6, the Association for Assessment in Counseling and Education (AACE), which is now the Association for Assessment and Research in Counseling, has published the *Standards for Multicultural Assessment* (AACE, 2012) that are included in Appendix C. In this document, it calls upon counselors to advocate for culturally diverse individuals from marginalized populations. This is particularly important in the area of career assessment where career choices may be eliminated because of cultural stereotypes. According to Evans (2013), the interests of individuals from oppressed groups, such as ethnic minorities, are often limited due to discrimination and socialization. She contended that underrepresented groups often gravitate toward "acceptable" occupations and careers,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

which restrict their career exploration. Evans (2013) also provides some suggestions for counseling strategies related to multicultural career assessment, which are the following:

- Review the assessment procedures with the client in order to ensure that the assessments will meet the client's needs.
- Investigate the assessment by reading all appropriate information (e.g., manuals, reviews, research articles) about the assessment's cultural inclusiveness, which should include information on language and reading level.
- Focus specifically on the standardization sample and ensure that the diversity percentages are close to the U.S. general population. If there are no ethnic minorities in the norming sample, then there needs to be a clear justification for using the instrument (e.g., the client is highly acculturated).
- If ethnic minorities are included in the norming sample, then check to see if any analysis of group differences has been conducted. If the sample of ethnic minorities is too small and/or there are no cultural comparison data, then the assessment also needs to be justified.
- Analyze comparative data regarding differences in cultural groups and determine whether or not to proceed with the assessment. See if there are norms for your client's ethnicity or cultural group. If there are testing differences and no norms for your client's cultural group, then proceed with caution and determine how to use the information in reporting the results to the client.
- Always explain to the client any caveats before administering a career assessment and reiterate these caveats again when interpreting the results.
- Provide the best interpretation possible and take into account the client's race, culture, and environment.

Summary

Testing and assessment has often been seen as an integral part of career counseling. This chapter reflects many of the common practices in the area of career assessment. When considering career assessment, one should not just consider using formal assessments in isolation; Flores, Spanerman, and Obasi (2003) contended that career assessments should be supplemented by interviewing so that counselors can gather a wide range of information. Turner et al. (2001) suggested that career/vocational counselors should be able to integrate comprehensive results that cross the multiple domains of adolescent and adult development, personality, and psychopathology with detailed knowledge related to the assessment of interests, abilities, personality dimensions, and values. Furthermore, Chartrand and Walsh (2001) asserted that personality and cognitive assessments should be considered part of the career assessment domain. Therefore, career assessment and counseling do not involve solely job-related issues; rather, career assessment and counseling concern viewing the individual holistically and integrating career and personal information.

A number of well-developed measures of clients' individual attributes can be used effectively in career assessment. In career counseling, counselors often assess clients' interests, which are a good indicator of career direction or choice. In terms of interest inventories, John Holland's theory has had a significant influence. One of the instruments influenced by Holland is the Strong Interest Inventory®, which is among the most widely used instruments in counseling. The purpose of the Strong is to identify interest patterns. The Strong Interest Inventory® provides results related to 6 General Occupational Themes, 30 Basic

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Interest Scales, and 244 Occupational Scales (122 for women and 122 for men). Counselors can have the most confidence in interpreting the General Occupational Themes (GOTs) and the Basic Interest Scales (BISs), and less confidence in the results from the Occupational Scales (OSs) and the Personal Style Scales. There are other interest inventories or methods for assessing that counselors can consider in the career counseling process.

Ability assessment or aptitude testing is often used in career counseling because they are significant predictors of occupational success. There is also research that supports the use of self-estimates of abilities. The exploration of clients' aptitudes and abilities should include an analysis of their self-efficacy beliefs. Because self-efficacy beliefs directly influence the tasks attempted by clients, they should be considered when clinicians are engaged in any therapeutic relationship. The assessment of values is sometimes neglected despite being an important part of the career assessment process. Work values have been found to be highly correlated with work satisfaction. Any values instrument is not inclusive of all values; hence, these assessments need to be supplemented with other exercises when working with clients. There are a number of integrated career assessment programs that involve the assessment of interest, abilities, and work values. Some of the more intensive programs also integrate occupational and education information into the programs so that the assessments and germane information are woven together to assist the client.

Rather than focusing on individual differences, some career assessment tools focus on where clients are in the career decision-making process. Typical measures of career process variables are related to the level of decidedness or certainty and degree or stage of career maturity or adaptability. These instruments can be helpful in selecting career interventions that are appropriate for the client. In addition, these instruments are used to evaluate career counseling to see if clients became more vocationally mature or decided about their occupational choice as a result of the counseling. Counselors may also want to consider qualitative career assessments, which have evolved from a postmodern or constructivist perspective. Qualitative career assessments focus on understanding the client's view with an understanding that a client's perspectives are influenced by cultural, societal, and relational factors. Qualitative career assessment often involves analyzing clients' discourse or stories to identify themes and perspectives. In conducting qualitative career assessments, it is important that counselors understand the philosophical foundations of these techniques.

Technology has had a significant influence on the area of career assessment, and it is anticipated that this trend will continue. Computers are very adept at storing large amounts of information, and some individuals saw that this capacity made computers an excellent tool for combining career assessment with the delivery of career information. The number of online career assessments has substantially increased in the past few years. These assessments include both the assessments developed by professionals and the quizzes and surveys developed by lay individuals. In terms of both Internet assessments and computerized systems, these instruments, like more traditional assessments, should be evaluated in terms of the reliability, validity, norms, and suitability for the client.

In conclusion, counselors involved in career assessment must be particularly cognizant of gender issues and of ethnic and racial differences on career instruments. These issues are being addressed in the field, and, in response, instruments are being developed and revised. Although instrument developers are attending to these issues, however, the practitioner still bears the ultimate responsibility for appropriate use of an instrument. Therefore, clinicians must ensure that career assessments are appropriate for each client they assist.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 12

Appraisal of Personality

Learning Objectives

- LO1 Distinguish between structured personality assessments and projective techniques
- LO2 Identify the strengths and limitations of assessing personality using observation and interviewing
- LO3 Understand the purposes of the validity scales of the MMPI-2
- LO4 Describe the major differences between the MMPI-2 and the MMPI-2-RF
- LO5 Identify the "big five" factors of personality as measured by the NEO-PI-3
- LO6 Describe the four dichotomies or dimensions measured by the Myers-Briggs Type Indicator
- LO7 Ascertain the strengths and limitations of projective techniques
- LO8 Recognize the issues in the assessment of self-concept or self-esteem

As a counselor, there is a good possibility that you will occasionally conduct workshops on topics related to counseling. For example, consider that you have been asked by a utilities company with a large number of employees to conduct a two-day workshop on reducing stress and increasing productivity for the clerical staff. As a part of the workshop, they would like for you to give the participants a personality inventory. The company's vice president for human resources believes a personality assessment will help the staff understand themselves better and could identify ways the staff could cope better with the stresses of their jobs. This organization has approximately 60 employees on the clerical staff, so you will need a personality inventory that can be used with a variety of people and that does not focus exclusively on psychopathology. You will also need to select an instrument that is easy to score, as it is only a two-day workshop.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L01 The term personality is used in everyday conversation and has certain connotations in the mental health professions. It is not uncommon to hear someone say, "She has a good personality." When we attempt to define personality, we find that it is a somewhat elusive construct. *Personality* comes from the Greek word *persona*, which was a role-play by an actor. Like other psychological constructs, there are varied theoretical views about the facets of personality. Recent research has contributed to a greater understanding of the structure of personality. Although we know more about personality than we did 20 years ago, there is still debate about the precise definition and how best to measure it.

The assessment of personality has existed for thousands of years. The early hunters and gatherers may have assessed each other's personalities to determine who would be a better hunter or who would be a better gatherer. Nonetheless, the field of formal psychological assessment of personality is less than a century old. During the twentieth century, both public sentiments and professional views toward personality assessment varied. Butcher (1995) predicted that personality assessment was going to increase, but Piotrowski's study (1999) found that with the rise of managed care, personality assessment had decreased in use. In the future, the popularity of personality assessments will probably wax and wane based on professional, political, and economic issues.

Incorporating personality assessment into the counseling process can be beneficial. Personality assessment can, at times, provide a shortcut in identifying client problems. In selecting interventions and making decisions about treatment, counselors can frequently benefit from having information about the client's personality. Personality probably influences coping styles, needs and desires, responses to environmental stressors, interpersonal patterns, and intrapersonal sensitivity. Counselors can use client personality information to assist them in structuring the counseling relationship and in selecting specific interventions for that client.

This chapter provides an overview of personality assessment. The information contained in this chapter is not sufficient for anyone to use these assessments without considerable additional training and supervision. Many of the instruments discussed in this chapter can be administered, scored, and/or interpreted by master's level practitioners; however, some formal instruments (e.g., Rorschach) cannot. Nonetheless, counselors often work in multidisciplinary teams and need to have a basic understanding of these instruments.

There are numerous informal and formal personality assessment tools. Within formal personality assessment, there are two major categories: **structured personality instruments** and **projective techniques**. Structured instruments (e.g., the Minnesota Multiphasic Personality Inventory-2 [MMPI-2]) are formalized assessments in which clients respond to a fixed set of questions or items. Projective techniques vary from standardized instruments in that the client is asked to describe, tell a story, or respond in some way to relatively unstructured stimuli. Weiner (2013) contended that a better term for *projective techniques* is *performance-based measures*, which is a broader term that includes all types of indirect measures of personality. I will continue to use the term *projective techniques* as it is more universally accepted. The intent of the personality assessment is less obvious with projective techniques than with a structured inventory. The unstructured format presumably guards against faking because the examinee is not sure what the examiner is evaluating. With projective techniques, clients are believed to "project" their personality characteristics in their responses. Thus, proponents of projective techniques consider them to be an effective method for uncovering latent, hidden, or unconscious aspects of personality.

This chapter examines both informal and formal personality assessment techniques. Informal assessment techniques are frequently used, and clinicians need to understand the reliability and validity of these measures. Informal techniques, just like formal instruments, need to be analyzed for their soundness. Following the discussion of the informal techniques

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

of observation and interviewing, there is a discussion of formal personality instruments. It includes an overview of the most commonly used structured personality inventories and projective techniques.

LO2 Informal Assessment Techniques

Observation

When you think of personality assessment, you may only consider formal assessment instruments such as the MMPI-2. If you believe personality assessment only includes standardized instruments, then you would be ignoring the most common method counselors use to assess personality: **observation**. When you meet someone, you may begin to make judgments about that person's personality based on such factors as attire, stance, choice of words, and facial expressions. In the same way, counselors observe clients from the first meeting and begin to make clinical judgments based on those initial observations. In fact, Blacher, Murray-Ward, and Uellendahl (2005) found that the most commonly used assessment tool reported by school counselors was observation. Counselor training is typically designed to produce individuals with refined observational skills. Social psychology research, however, has found that mental health practitioners tend to be biased when interpreting their observations. When clinical decisions are based on observation, there are problems with selective recall, selective interpretation, and the power of preexisting assumptions (Dawes, 1994). Clinicians often selectively remember certain aspects of what clients say and how clients behave. Practitioners also have "themes" concerning what they remember, and they have been found to recall similar information about a number of different clients. In addition, clinicians tend to be selective in their interpretations and to focus on certain behaviors and draw conclusions based on those behaviors. Social psychological research has also shown that people tend to see things with which they are familiar. Hence, a counselor who works primarily with clients who have eating disorders is more likely to see a new client as having an eating disorder than is a clinician who does not work with this type of client (Dawes, 1994). Subjectivity is a persistent problem in observation. As an example, you have probably experienced subjectivity in observation when a friend or family member holds a very different view of an event that you both witnessed.

Practitioners cannot afford to be subjective in their observations when assessing personality. Observation is an assessment tool; therefore, it should be evaluated in the same manner that we evaluate other assessment tools (e.g., examining the reliability and validity of our observations). Let us consider reliability in terms of observation. *Random error* has a way of slithering into counselors' observations of their clients. Counselors, like other humans, tend to get an occasional cold, get stuck in traffic, or have a disagreement with a family member. Any of these situations may lead to random error in their observations. Furthermore, the client may contribute to the random error. A counselor may perceive a client with sluggish speech and frequent yawns as being disinterested. However, these actions may be more related to the client staying up late the night before than to disinterest. Counselors should examine the reliability of their observational methods, for there may be a high degree of error in these methods. Counselors can adapt some of the traditional methods of determining reliability to examine the consistency of their own observations. For example, counselors could systematically conduct a test-retest situation by noting observations in one session and then comparing these with recorded observations of a second session. A counselor could split a session into halves or systematically attend to observations at specific time intervals in a session to examine the consistency of these observations.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

There also can be problems with the validity of counselor observations. When counselors observe, they need to consider the *representativeness* and the *generalizability* of their observation. In terms of representativeness, because the time the client spends with the counselor is not long, the client may not be completely natural with the counselor. Therefore, the validity of the observations may be restricted because only a small sample of behaviors was observed, which might not represent the client's typical behavior. Inadequate observations limit the counselor's ability to generalize as to how the client behaves outside of counseling. For example, a client in the counseling sessions may act reserved; yet, when outside of counseling, the client's behavior may be very domineering and aggressive. One method for enhancing the representativeness of the observations is to increase the amount of time the individual is observed. For school counselors, this might be easily accomplished by observing students in the classroom, during lunch, during physical education, or on the playground. In other settings, counselors may want to have clients keep a log, which is essentially a self-observation technique. Another technique to increase the sample of observation is to have another person (e.g., a family member) record certain behaviors.

In conclusion, counselors often use observation to assess personality, and in doing so, they need to consider certain factors. Clinicians need to consider carefully how they organize the observations as well as the conceptual framework they use in determining their conclusions. Furthermore, clinicians should evaluate their working models of personality that may influence this process. Counselors also need to consider how their observations are filtered through their own biases and beliefs.

Interviewing

Interviewing is another widely used method of personality assessment. In this context, an *interview* is defined as a face-to-face verbal exchange in which the counselor is requesting information or expression from the client. Vacc and Juhnke (1997) found that structured clinical interviews fall into the categories of either *diagnostic assessments* or *descriptive assessments*. The intent of diagnostic interviews is to identify issues and possible disorders consistent with a diagnostic taxonomy (e.g., the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition*). Chapter 15 addresses interviews in terms of diagnosis. Descriptive interviews are used when the purpose is to describe aspects of the client. The focus in this chapter is on descriptive interviews with the specific purpose of describing personality.

The same characteristics of a good intake interview (Chapter 8), such as being warm and encouraging, also apply here. The quality of the questions that counselors use in assessing personality has a substantial influence on the value of the assessment. Poorly developed questions that are not conceptually related to personality assessment will probably add little to the counselor's understanding of the client's personality. Sometimes, counselors will use questions related to the client's problems and situation as the primary mechanisms for personality assessment. Questions only peripherally related to personality may not provide clients with the opportunity to explain aspects of their personality. Therefore, when the major purpose is to understand the client's personality, counselors should ask clients questions that directly assess personality. As we will see later in this chapter, there is substantial research supporting a five-factor model of personality, and a structured interview has been developed to assess those factors of personality. The *Structured Interview for the Five-Factor Model of Personality* (Trull & Widiger, 1997) can involve more than 120 questions as the interview frequently requires follow-up questions to clarify the client's responses. In concluding this section on assessing client's personality based on interviewing, it is important to note that many of the same issues related to the discussion of reliability and validity of observation also applies to interviewing.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Structured Personality Inventories

To understand the most commonly used personality instruments in counseling, it is probably valuable to have some foundational information regarding the standard methods used in constructing personality inventories. Anastasi and Urbina (1997) suggested that there are four basic methods of constructing personality inventories. The first method is the *content-related procedure*, in which the focus is on the content relevance of the items. Developers using this method follow a rational approach to item development, with items directly relating to the personality attributes being measured. The content scales of the MMPI-2 are examples of this type of procedure. There are some difficulties with content-related procedures because the instruments are often easy to fake or to have a response set. A response set exists when an individual answers one question a certain way and then falls into the habit of answering similar questions in the same way.

The second method of constructing personality inventories uses *personality theory* as the base for development. Once the instrument is developed, construct validation procedures are implemented to determine if the instrument actually measures the tenets of the theory. Thus, the validation information is interpreted concomitantly with the theory. An example of a theory-related instrument is the Myers-Briggs Type Indicator, which is based on Jungian theory.

Empirical criterion keying is the third method of constructing personality inventories. Items are selected based on their relationship to some external criterion rather than on their content. The MMPI-2 is an example of the empirical criterion keying method, because items were selected that separated people who were considered normal from those who were diagnosed with some form of psychopathology. The Occupational Scales of the Strong Interest Inventory® were also developed using this same method—only items that differentiate between people working in an occupation from people not working in that occupation are used in scoring each occupation scale.

The last method involves the statistical technique of *factor analysis*. This strategy entails examining the interrelationships of items and determining the similarities of the items that group together. Researchers who have used factor analysis to investigate personality have found five factors (nicknamed the "Big Five"). The NEO Personality Inventory—3 (NEO-PI-3) is designed to measure these five factors of personality.

The precise number of self-report personality inventories is difficult to determine, but there are at least several hundred. Rather than discuss all of these instruments (which would mean reading this book for a very long time), this section summarizes the instruments most frequently used by counselors. The most widely used structured personality instrument, the MMPI-2, will be discussed first. The second instrument to be discussed is the NEO-PI-3, which has a substantial empirical foundation and is increasingly being used by practitioners. Finally, the Myers-Briggs Type Indicator, another popular instrument used in diverse settings, will also be reviewed.

LO3 Minnesota Multiphasic Personality Inventory-2

The *Minnesota Multiphasic Personality Inventory-2* (Butcher, Dahlstrom, Graham, Tellegen, & Kraemmer, 1989) replaced the original MMPI in order to improve on this widely researched and used instrument. The MMPI-2 subsequently replaced the MMPI as the most widely used personality instrument. A revised version of the MMPI-2 manual was published by Butcher et al. (2001). Soon after the publication of the MMPI-2, another form was introduced, the *Minnesota Multiphasic Personality Inventory—Adolescent Version* (MMPI-A; Butcher et al., 1992) for children ages 14 to 18. The MMPI and the MMPI-2 are empirical criterion keyed instruments, which means the items were not selected for their content but rather for their relationship to a criterion. The criterion in the original MMPI

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

was the identification of psychopathology: it was designed to differentiate those individuals with psychopathology as compared with normal individuals. The MMPI-2 manual reflected that it is a "broad-based test to assess a number of the major patterns of personality and emotional disorders" (Butcher et al., 1989, p. 2). Therefore, the MMPI-2 is used to diagnose emotional disorders, but it is also intended for use in nondiagnostic activities. Although the MMPI-2 is often cited as the most widely used structure personality assessment, Maruish (2004) has noted a decline in use because of its length (i.e., 567 items). On the other hand, the MMPI-2 continues to be actively researched as evidenced by Graham's (2006) finding that, since the publication of the MMPI-2 in 1989, there have been approximately 2,800 published journal articles related to the MMPI-2.

The MMPI-2 booklet contains 567 items, and the MMPI-A has 478 items. The items are statements (e.g., I am sure I get a raw deal from life; I cry easily) to which the client responds by indicating whether the statement is true, false, or cannot say. The MMPI-2 has an eighth-grade reading level, although clinicians can purchase an audiocassette version for clients whose reading level is not sufficient. In addition, the MMPI-2 is available in many other languages.

One of the major reasons for revising the MMPI was that the original norming group had been tested in the 1930s and consisted primarily of white, middle-class individuals from Minnesota (Butcher, 2011). The MMPI-2 norming group, selected to match the 1980 census data, included 2,600 individuals who matched the proportions of ethnic groups, geographic representation, and age distribution of the U.S. population. The MMPI-2 norming group has been criticized, however, for including a high proportion of individuals with advanced educational backgrounds and higher-status occupations. Although there is greater representation of minority individuals in the MMPI-2 than in the MMPI, there continues to be some debate about racial bias (Newmark & McCord, 1996).

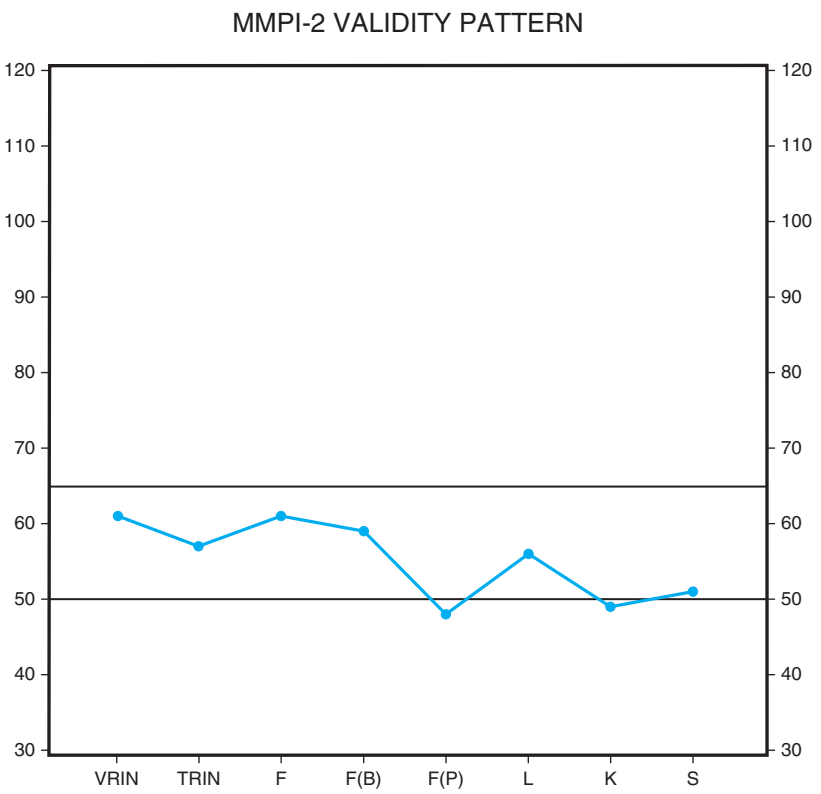
The interpretation of the MMPI-2 is complex, and specific training is needed to use the results appropriately. There are seven validity scales and a variety of clinical scales subdivided into three major categories: Basic Scales (10), Content Scales (15), and Special Scales (varying number) (MacCluskie, Welfel, & Toman, 2002). Butcher (2011) made the point the MMPI-2 was not designed to measure all characteristics or behaviors that may be of interest to the counselor. A proper interpretation involves examining the entire profile as a whole, not just each clinical scale in isolation. The MMPI-2 could be dangerous in the hands of the casual user who has not mastered its intricacies. Even though a *T* score of 65 is usually considered the cutoff, a clinician does not consider just one elevated score to indicate the disorder with which that scale is labeled. This is one reason that the basic clinical scales are usually referred to by number rather than by name. For example, an elevated Scale 4 (Psychopathic Deviance) might result from family discord, friction with peers, hostility toward a supervisor, or alienation from self and society. As Groth-Marnat (2009b) indicated, a clinician should not interpret an elevated Scale 4 score as always indicating antisocial acting out, as some groups tend to have somewhat elevated Scale 4 scores (e.g., counterculture groups, African Americans, graduate students in humanities and social sciences).

The purpose of this discussion of the MMPI-2 and MMPI-A is to provide counselors with an introduction to these well-known personality assessments. This discussion, however, does not provide sufficient information on how to interpret or use the MMPI-2 or the MMPI-A, and clinicians would need additional training and supervision before they would be qualified to use these assessments. In analyzing a client's results, one of the initial steps often involves examining how long the individual took to complete the MMPI and whether the client's time is close to the typical time (90 minutes for the MMPI-2 and 60 minutes for the MMPI-A). The clinician should have some concerns if the individual took an extraordinarily long time or completed it very quickly. A common second step involves

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

scoring and plotting the scores on the profile sheet or examining a computer-scored profile. Figure 12.1 MMPI-2 profile includes the validity scales. There are additional profile sheets for plotting the content scales and the supplementary scales. In addition, some MMPI-2 or MMPI-A reports will include a Welsh Code, which is a shorthand method for indicating performance on the Basic Scales (see Butcher & Williams, 2000). The clinician would then move on to examining the "validity" of the profile using primarily the validity scales (see the following section). The scores on the validity scales provide an indication of the degree to which the individual may have attempted to distort or skew his or her responses. After the validity scales are examined to determine the amount of distortion, the Basic Scales (which are described later in this chapter) are interpreted. Often when two or three Basic Scales are elevated, the interpretation of the profile includes an examination of the Two-Point Codes and Three-Point Codes. Many times, interpretations of the MMPI-2 will also include analysis of the 15 Content Scales, which are listed in Table 12.2.

FIGURE 12.1
MMPI-2 Basic Service Profile MMPI-2 Basic Service Report. Copyright © 1989 Regents of the University of Minnesota. All rights reserved. Portions excerpted from the MMPI-2 Manual for Administration, Scoring, and Interpretation. Revised Edition, copyright © 2001 Regents of the University of Minnesota. All rights reserved. "MMPI-2" and "Minnesota Multiphasic Personality Inventory-2" are trademarks of the University of Minnesota



	VRIN	TRIN	F	F(B)	F(P)	L	K	S
Raw Score:	8	8	8	4	1	5	15	26
T Score:	61	57F	61	59	48	56	49	51
Response %:	100	100	100	100	100	100	100	100

Cannot Say (Raw):	0
Percent True:	37
Percent False:	63

	Raw Score	T Score	Resp. %
S1 - Beliefs in Human Goodness	5	44	100
S2 - Serenity	7	53	100
S3 - Contentment with Life	4	50	100
S4 - Patience/Denial of Irritability	7	63	100
S5 - Denial of Moral Flaws	2	43	100

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Validity scales. In interpreting a personality inventory, it is often quite helpful to understand individuals' attitudes while they were taking the inventory. For example, it might be helpful to know if clients are trying to put their "best foot forward," or if someone on trial for murder is attempting to appear psychologically disturbed. The validity scales of the MMPI-2 are designed to provide this type of information; in other words, these scales reflect the validity of the profile. There are several measures of the validity of the MMPI-2 and MMPI-A profile, and it is best to examine many of the measures simultaneously to make clinical judgments about the usefulness of the results. Butcher and Williams (2000) recommended first examining item omissions, inconsistent responding, and fixed responding, which involves the Cannot Say Score (?), VRIN, and TRIN. They then suggested that a clinician examine measures of random and exaggerated responding, which are the F, Fb, Fp, and FBS scales. Finally, the clinician should examine the scales designed to assess defensiveness and claims of extreme virtue, the L, K, and S scales.

Cannot Say Score (?) The Cannot Say Score (?) Scale is simply the total number of items on the MMPI-2 that the individual did not answer. Butcher (2011) suggested examining if there is a pattern to the omissions. If a client leaves a large number of items blank on the MMPI-2 or MMPI-A, it is important to explore possible factors that may have contributed to the omissions (e.g., difficulty with reading, confusion, recalcitrance). In general, if an individual leaves 30 or more questions unanswered, then the profile is considered invalidated (i.e., too many omissions to be interpreted).

VRIN Scale. The Variable Response Inconsistency (VRIN) Scale examines consistency of response and can be helpful in determining if the person randomly marked answers or had difficulty understanding the items. VRIN consists of paired items in which the content is very similar or opposite. Thus, if responses are not in the expected manner, then the inconsistencies will be reflected in the VRIN score. VRIN can be helpful in interpreting whether a higher F Scale relates to responding randomly or attempting to fake.

TRIN Scale. The True Response Inconsistency (TRIN) Scale is designed to measure "yea-saying" or "nay-saying." These are paired items with consistency indicated by answering true one time and false the other time. This scale identifies those individuals who mark all the items true or all the items false.

F Scale. The Infrequency (F) Scale concerns whether the individual is faking or attempting to exaggerate symptoms. Endorsing a large number of these items indicates the test taker is presenting an extremely symptomatic picture not found in the general population. In some cases, there may be a tendency to exaggerate symptoms or attempt to "fake bad." A score of 90 is considered to reflect exaggerated negative symptom presentation (Butcher, 2011).

F(B) Scale. The Infrequency Back [F(B)] Scale is an extension of the F Scale and measures items that are infrequently endorsed by the general population. The majority of validity scales are in the first part of the MMPI-2, in case practitioners use the shortened version; the FB is on the back portion of the MMPI-2 and can verify the score on the F Scale. In addition, counselors may also use the score on the FB to determine whether the person may have responded randomly on the second half of the test. The MMPI-2, which has 567 items, is a lengthy instrument, and some test takers may become tired or bored while completing the second half of the assessment.

F(p) Scale. The Psychopathology Infrequency [F(p)] Scale is a comparatively new validity scale that indicates the veracity of the client's negative symptoms. This validity scale is

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

designed to indicate rare or extreme responses in a psychiatric setting as compared with the other F scales, which indicate rare responses in a normal setting. This scale was designed to address possible malingering of psychological symptoms in a mental health treatment context (Butcher, 2011).

FBS Scale. The Symptom Validity (FBS) Scale has been recently added to the standard validity scales of the MMPI-2, although it was introduced by Lees-Haley, English, and Glenn (1991). It is sometimes informally labeled the fake bad scale, but there is substantial research that supports this scale's usefulness in measuring potentially exaggerated claims of disability. This scale should not be interpreted in isolation: it should be interpreted in the context of the setting and in conjunction with the scores of the other validity scales. There is some controversy about the FBS (see Butcher, 2011).

L Scale. The Lie (L) Scale provides an indication of the degree to which the individual is trying to look good. Higher scores indicate that the individual wants to report socially correct behaviors; lower scores suggest the ability to acknowledge general human weaknesses. This scale should not be interpreted to represent overall honesty or the tendency to lie. A high L Scale may indicate that the counselor will need to work on being seen as trustworthy to encourage the client to be more self-disclosing as it is a good indicator of underreporting of symptoms.

K Scale. The Correction (K) Scale measures defensiveness or guardedness. It is more subtle than the L Scale but measures the same dimension of trying to present oneself in an overly positive manner. The context in which the person is taking the MMPI-2 must be considered when interpreting the K Scale. For example, one might expect a higher K Scale if the MMPI-2 is a part of a hiring process than if it is taken by someone seeking counseling. In counseling situations, an elevated K Scale (above 65) may indicate the individual will have difficulty recognizing his or her own contributions to problems.

S Scale. The Superlative Self-Presentation (S) Scale is another comparatively new validity scale of the MMPI-2 and was designed as an extra measure of defensiveness. In addition to serving as a general measure of defensiveness, it is also possible to obtain subscale scores that provide information on the possible reasons underlying the defensive attitude. These subscales are the Beliefs in Human Goodness, Serenity, Contentment with Life, Patience/Denial of Irritability and Anger, and Denial of Moral Flaws.

Basic or Clinical Scales. After examining the validity scales to determine the validity of the profile, the next typical step is to interpret the basic or clinical scales of the MMPI-2 or MMPI-A. These scales have served as the heart of the MMPI and have generated thousands of books and articles (Butcher, 2011). Table 12.1 lists the 10 scales, each with a number, an abbreviation, and a formal name. Clinicians generally refer to the scale number rather than the formal name of the scale. As with the validity scales, the clinical scales are not interpreted in isolation because other high scores are taken into account when interpreting the MMPI-2. In MMPI-2 interpretation, the term *elevated* is often used to indicate that the *T* score on that scale is above 65. An interpretation may also include the term *moderately elevated*, which generally indicates that the *T* score is between 60 and 65. Sometimes a practitioner focuses only on the scales with elevated scores; however, other scales should also be examined to identify possible strengths and coping skills. The following overview of each clinical scale is designed to provide a basic understanding of the scales and an introduction to how the scales are combined in interpretation.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 12.1
The Clinical or Basic Scales of the MMPI-2

Number	Abbreviation	Formal Name
Scale 1	Hs	Hypochondriasis
Scale 2	D	Depression
Scale 3	Hy	Conversion Hysteria
Scale 4	Pd	Psychopathic Deviate
Scale 5	Mf	Masculinity-Femininity
Scale 6	Pa	Paranoia
Scale 7	Pt	Psychasthenia
Scale 8	Sc	Schizophrenia
Scale 9	Ma	Hypomania
Scale 0	Si	Social Introversion

Scale 1: Hypochondriasis. Scale 1 is related to excessive concerns with health, as indicated by reporting the presence of bodily complaints. This scale assumes that there is little or no organic basis for these complaints. If you have a client with a chronic illness or a client who was physically ill while taking the MMPI-2, then you need to be more cautious in your interpretation. It is believed, however, that scale 1 represents a characterological preoccupation with bodily concerns rather than a situational attentiveness to physical complaints. In addition to the physical complaints, elevated scores indicate a tendency to be self-centered, selfish, and pessimistic—although the complaints may center on health issues, they can spread to complaining or whining about other issues. Markedly elevated scores may indicate the tendency to manipulate others with physical complaints and to use these complaints to avoid responsibility and meet dependency needs.

Scale 2: Depression. Elevated scores on scale 2 suggest feelings of discouragement, hopelessness, and isolation. Lower scores on the scale indicate the tendency to be optimistic, gregarious, and alert. As the scores on scale 2 increase, the client's attitude changes from sadness or gloom to a pervasive pessimism about the world. This pessimism extends to the self, with high scale 2 scores often indicating self-deprecation and overriding feelings of guilt and hopelessness. Individuals with high scale 2 scores tend to lack self-confidence and worry over small things.

Scale 3: Conversion Hysteria. Scale 3 concerns conversion disorders that are a complex combination of denial and social flamboyancy. Moderate elevation may only reflect a tendency to think positively about people. Elevated scores reflect the tendency to deny issues. Individuals with elevated scores often react to stress and anxiety with physical symptoms. Elevations on scale 3 are also indicative of individuals who are extroverted but somewhat superficial. They have a tendency to lack insight into their own motivations and actions as well as insight into others' motivations. Individuals with a high scale 3 score tend to expect a great deal of attention and affection from others. Rarely is a person with an elevated scale 3 score diagnosed as psychotic.

These first three scales of the MMPI-2 are sometimes examined in combination. Sometimes they are labeled the "neurotic triad," which is a somewhat misleading title. The configurations among the triad of scales are examined. There is empirical support for certain interpretations based on certain patterns.

Scale 4: Psychopathic Deviate. An elevation on scale 4 suggests that the client is "fighting something," but does not identify the specific conflict. Higher scores reflect hostility

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

toward social convention, authority figures, and possibly family members. Descriptors of individuals with elevated scale 4 scores include moody, irritable, resentful, and impulsive. These individuals tend to blame other people for their troubles. In addition, they often have difficulties with law enforcement and may be incarcerated.

Scale 5: Masculinity-Femininity. Scale 5 provides a measure of whether the individual's emotions, interests, attitudes, and perspectives on social relationships and work are consistent with men or women in general. This scale is scored separately—men's attitudes are compared with men and women's with women. High scores indicate that a person's general perspective is nontraditional from others of the same gender. Hence, for males, elevated scores are associated with having a wide range of interests, especially aesthetic interests. On the other side, low scores for males indicate more of a "he-man" attitude. Higher scores are also associated with nontraditional attitudes for women, with elevated scores indicating interest in masculine activities, being assertive, competitiveness, and self-confidence. Women with lower scores have endorsed more items indicative of stereotypically feminine interests. Low scores for women on scale 5 may indicate passivity and submissiveness. Any interpretation of scale 5 needs to incorporate the client's ethnicity, socioeconomic status, and educational level.

Scale 6: Paranoia. Scale 6 considers whether there is a pattern of suspiciousness, mistrust, delusional beliefs, and marked interpersonal sensitivities. Some of the items also address self-centeredness and insecurity. Moderately elevated scores are characteristic of a sensitive individual. An elevated scale 6 score indicates the individual has the tendency to misinterpret the motives and intentions of others. These individuals tend to be suspicious and believe that others' actions are somehow aimed at them. Markedly elevated scores are indicative of disturbed thinking and psychotic behaviors.

Scale 7: Psychasthenia. Scale 7 is related to some of the symptoms associated with anxiety and obsessive-compulsive disorders. It is often considered to be a measure of neuroticism and generalized anxiety and distress. Moderate scores may reflect the ability to be methodical, organized, punctual, and have high moral standards. Elevated scores indicate a tendency toward self-blame accompanied by anxiety, rumination, and feelings of insecurity. Extremely high scores are usually associated with some obsession and rigid efforts to control impulses.

Scale 8: Schizophrenia. The developers of the original MMPI constructed scale 8 to identify patients manifesting various forms of schizophrenia. Because various forms of schizophrenia are measured with this scale, there are a large number of items that may appear somewhat strange to more typical clients. Scale 8 includes a wide range of strange beliefs, unusual experiences, and emotional problems. Elevated scores reflect an unconventional lifestyle, feelings of alienation, and psychoticism. It may be difficult to follow the individual's train of thought and mental processes. High scores are usually associated with people who are isolated and alienated. Extremely high scores are associated with individuals who show blatantly psychotic symptoms.

Scale 9: Hypomania. Scale 9 was constructed to measure manic or hypomanic behaviors, such as being energetic, hyperactive, and euphoric. Moderately elevated scores (not uncommon in graduate school students) suggest highly energetic, involved, and gregarious people. As the scores increase, individuals are expending more energy but are less efficient ("spinning their wheels"). Elevated scores are indicative of overactivity, flight of ideas, and emotional excitement. Clients can be easily frustrated and may fluctuate between euphoria and irritability.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Scale 0: Social Introversion. High scores on scale 0 are associated with a preference for being alone; low scores indicate a preference for being with others. With many of the scales of the MMPI-2 and MMPI-A, clinicians have a tendency to assume that low scores are positive and high scores are negative. This conclusion on scale 0 is incorrect because introversion is not necessarily a deficit. Clients with moderately elevated scores on this scale prefer to be alone or with select friends. Increased scores usually indicate more social withdrawal and anxiety associated with being around people. Individuals with high scores have a tendency to worry and be guilt-prone. Low scores are usually not interpreted with many of the clinical scales of the MMPI-2, but that is not the case for scale 0. Low scores indicate the person is outgoing and gregarious; however, very low scores may be associated with a tendency to be immature, impulsive, and self-indulgent.

Two-Point Codes and Three-Point Codes. After examining the client's scores on each of the basic scales, an interpretation of the MMPI-2 will often include an analysis of the two or three elevated scales that draw from the research on code types (Butcher & Williams, 2000). Butcher and Williams, however, cautioned against using code types when interpreting MMPI-A results. Two-point code types are one of the more researched areas of the MMPI-2 and occur when two basic scales scores are in the critical range (*T* scores at or above 65). For example, if an individual's scales 1 and 9 scores were 66 and 70, respectively, this would be a two-point code of 1-9/9-1. Butcher and Williams provided information on 20 different two-point code types and 3 common three-point code types. These descriptions include symptoms and behaviors, personality characteristics, and prediction or dispositions.

Other Supplemental Scales. Since the original MMPI was published in the 1940s, many other scales have been developed to supplement the original validity scales and clinical scales. The content scales have been one of the more widely researched and used scales after the basic or clinical scales of the MMPI-2. As compared with the clinical scales, which were empirically developed, the content scales were developed by examining item content. Barthlow, Graham, Ben-Porath, and McNulty (1999) provided evidence that the content scales significantly add to predicting clients' behavior and personality, even more so than the clinical scales. The 15 content scales are briefly described in Table 12.2. On the MMPI-A, 11 of the content scales are the same as the MMPI-2; however, the scales of Fears, Antisocial Practices, Type A Personality, and Work Interference are replaced by the scales of Alienation, Conduct Problems, Low Aspirations, and School Problems.

The number of supplementary scales that can be derived from the MMPI-2 is substantial. Some of the supplementary scales can be hand scored; because of the time investment, however, most practitioners use computer-generated supplementary scales. Clinicians use these additional scales to both augment and confirm their interpretation of the clinical scales. Most of the supplemental scales were developed with a specific purpose in mind (e.g., addiction problems, marital distress). The supplemental scales that Butcher (2011) endorsed as providing the clinically useful information are the MacAndrew Alcoholism Scale (MAC-R), the Addiction Potential Scale (APS), the Addiction Admission Scale (AAS), the Marital Distress Scale (MDS), the Hostility (Ho) Scale, and Posttraumatic Stress Disorder Scale. He also suggested that clinician use the Psychopathology Five (PSY-5) scales (i.e., Aggressiveness, Psychoticism, Disconstraint, Negative Emotionality/Neuroticism, and Introversion/Low Positive Emotionality). The Psychopathology Five (PSY-5) scales are designed to help in the diagnosis of personality disorders.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 12.2
The Content Scales of the MMPI-2

Content Scale	Descriptors/Symptoms
Anxiety (ANX)	Tension, somatic problems, sleep difficulties, excessive worry, and concentration problems
Fear (FRS)	Specific fears or phobias (excluding general anxiety)
Obsessiveness (OBS)	Rumination and obsessive thinking, difficulties with decisions, and distress with change
Depression (DEP)	Depression, feeling blue, uninterested in life, brooding, unhappiness, hopelessness, frequent crying, and feeling distant from others
Health Concerns (HEA)	Many physical complaints across body systems, worry about health, and reports of being ill
Bizarre Mentation (BIZ)	Thought disorder that may include hallucinations, paranoid ideation, and delusions
Anger (ANG)	Anger control problems, irritability, being hotheaded, and having been physically abusive
Cynicism (CYN)	Misanthropic beliefs, suspicion of others' motives, and distrustful of others
Antisocial Practices (ASP)	Misanthropic attitudes, problem behaviors in school, antisocial practices, and enjoyment of criminals' antics
Type A Personality (TPA)	Hard-driven and competitive personality, work-oriented, often irritable and annoyed, overbearing in relationships
Low Self-Esteem (LSE)	Negative view of self that does not include depression and anxiety, feeling unimportant and disliked
Social Discomfort (SOD)	Uneasiness in social situations and preference to be alone
Family Problems (FAM)	Family discord, families seen as unloving, quarrelsome, and unpleasant
Work Interference (WRK)	Problems with and negative attitudes toward work or achievement
Negative Treatment Indicators (TRT)	Negative attitude toward physicians and mental health professionals

Psychometric Qualities. There has been some debate concerning whether the MMPI-2 is equivalent to the MMPI and, as such, whether it was appropriate to build from the vast amount of validation information that existed on the original assessment. In their reviews of the MMPI-2, both Archer (1992) and Nichols (1992) contended that, for better or worse, the instruments are closely related. Certainly, Butcher et al. (1989) provided a compelling case for the comparability of the two instruments. The test-retest reliability coefficients for the clinical scales of the MMPI-2 range between .67 and .92 for men and .58 and .91 for women. Using reliability generalization to analyze MMPI-2 coefficients, Vacha-Haase, Tani, Kogan, Woodall, and Thompson, (2001) found the average reliability coefficients ranged from .64 for scale 6 to .81 for scale 0. There is substantial validity information concerning the MMPI-2 and, to a lesser extent, the MMPI-A. Archer and Krishnamurthy's (1996) major concern about the MMPI-A is the instrument's ability to detect psychopathology and the possibilities of false negatives (i.e., the instrument indicates clients do not have a psychopathology when they actually do have a disorder). Austin (1994) criticized the MMPI-2 for not responding to recent developments concerning conceptual classification in psychopathology. Chapter 15 addresses diagnosis and some instruments (e.g., Millon Clinical Multiaxial Inventory) that are directly tied to a prominent diagnostic system.

In conclusion, the MMPI-2 and the MMPI-A are widely used instruments that have clinical applications for counselors. There are computer-generated reports of the MMPI-2 for various populations (e.g., outpatients, inpatients, college students). A counselor must have in-depth knowledge about this complex instrument to use it appropriately. Computer-generated interpretative reports have almost eliminated hand-scored and interpreted reports. The use of a computer-generated report, however, does not exonerate a clinician

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

from understanding the instrument. Furthermore, knowledge of the MMPI-2 does not mean someone can automatically use the MMPI-A, as there are differences. Ethical guidelines clearly indicate that knowledge of the specific instrument is needed.

L04 MMPI-2 Restructured Form

Sometimes there is some confusion between the MMPI-2 and the MMPI-2 Restructured Form (MMPI-2-RF), but they are separate instruments with different scales and are rooted in different research backgrounds (Butcher, 2011). Tellegen and Ben-Porath (2008) developed the MMPI-2-RF and contended it is a broad-spectrum inventory that is a construct-oriented revision of the MMPI-2. There are 338 items on the MMPI-2-RF that come directly from the MMPI-2. These 338 items measure 50 scales: 8 validity scales and 42 substantive scales (see Table 12.3). There is also indication that there will soon be a restructured form of the MMPI-A.

TABLE 12.3
Scales of the MMPI-2 RF

Validity Scales	
VRIN-r	Variable Response Inconsistency-Revised
TRIN-r	True Response Inconsistency-Revised
F-r	Infrequency-Revised
Fp-r	Infrequency-Psychopathology-Revised
Fs	Infrequency-Somatic
FBS-r	Fake Bad Scale-Revised
L-r	Lie-Revised
K-r	Defensiveness-Revised
Higher-Order Scales	
EID	Emotional/Internalizing Dysfunction
THD	Thought Dysfunction
BXD	Behavioral/Externalizing Dysfunction
Restructured Clinical (RC) Scales	
RCd	Demoralization
RC1	Somatic Complaints
RC2	Low Positive Emotions
RC3	Cynicism
RC4	Antisocial Behavior
RC6	Ideas of Persecution
RC7	Dysfunctional Negative Emotions
RC8	Aberrant Experiences
RC9	Hypomanic Activation
Somatic Scales	
HPC	Head Pain Complaints
NUC	Neurological Complaints
GIC	Gastro-Intestinal Complaints
Interest Scales	
AES	Aesthetic-Literary Interests
MEC	Mechanical-Physical Interests
Internalizing Scales	
SUI	Suicidal/Death Ideation
HLP	Helplessness/Hopelessness
SFD	Self-Doubt
NFC	Inefficacy
COG	Cognitive Complaints
SNV	Sensitivity/Vulnerability
STW	Stress/Worry
AXY	Anxiety
ANP	Anger Proneness
BRF	Behavior-Restricting Fears
MSF	Multiple Specific Fears
Externalizing Scales	
JCP	Juvenile Conduct Problems
SUB	Substance Abuse
AGG	Aggression
ACT	Activation
Interpersonal Scales	
FML	Family Problems
IPP	Interpersonal Passivity
SAV	Social Avoidance
SHY	Shyness
DSF	Disaffilitativeness
Personality Psychopathology Five (PSY-5) Scales	
AGGR-r	Aggressiveness-Revised
PSYC-r	Psychoticism-Revised
DISC-r	Disconstraint-Revised
NEGE-r	Negative Emotionality/ Neuroticism-Revised
INTR-r	Introversion/Low Profile Emotionality-Revised

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Similar to the MMPI-2, the MMPI-2-RF has validity scales that indicated the level of confidence the clinician can have in the results. Seven of the validity scales are revised from the MMPI-2 and one is new. The validity scales are the VRIN-r (Variable Response Inconsistency), TRIN-r (True Response Inconsistency), F-r (Infrequent Responses), Fp-r (Infrequent Psychopathology Responses), Fs (Infrequent Somatic Responses), FBS-r (Symptom Validity), L-r (Uncommon Virtues), and K-r (Adjustment Validity). There is also a Cannot Say indicator, and the computer-scored interpretation provides the percentage of questions that were answered true. Ben-Porath and Tellegen (2008) recommended that at least 90% of the items for any scale must be answered for interpretation of that scale.

The substantive scales are composed of the 9 Restructured Clinical (RC) Scales and 33 new scales. Tellegen et al. (2003) dramatically moved the interpretation of the MMPI-2 away from empirically keyed scales and completely revamped the building of clinical scales based on combining factor-analytic methods with construct-oriented scale development. Tellegen and colleagues called these new scales the Restructured Clinical Scales, and these scales have garnered significant clinical attention. These researchers noted that the scales of the MMPI were highly correlated with each other and lacked a theoretical base. They also criticized the empirical keying criterion as some of the criterion groups for the MMPI were often quite small. Tellegen et al. hypothesized that a "general distress" or demoralization factor was underlying a number of the clinical scales that obscured the interpretation of the MMPI-2 clinical scales. The overall goals in developing the RC Scales were to (a) have nonoverlapping scales, (b) identify scales that were distinct from Demoralization, and (c) identify scales that are relevant to clinical constructs (Rogers, Sewell, Harrison, & Jordan, 2006). The RC scales were developed to provide a roadmap for interpreting the MMPI-2 by measuring more precisely focused and pertinent elements of the clinical or basic scales (Tellegen & Ben-Porath, 2008). The RC Scales are currently included in the MMPI-2 Extended Score Report and the MMPI-2-RF reports. The RC Scales are RCd Demoralization, RC1 Somatic Complaints, RC2 Low Positive Emotion, RC3 Cynicism, RC4 Antisocial Behavior, RC6 Ideas of Persecution, RC7 Dysfunctional Negative Emotions, RC8 Aberrant Experiences, and RC9 Hypomanic Activation (see Table 12.3).

In developing the MMPI-2-RF, Tellegen and Ben-Porath (2008) set out to augment the RC Scales with a set of higher-order and more narrowly focused measures. As reflected in Table 12.3, the three Higher-Order Scales are Emotional/Internalizing Dysfunction (EID), Thought Dysfunction (THD), and Behavioral/Externalizing Dysfunction (BXD). Tellegen and Ben-Porath's further goal was to extract more problem-specific clinically relevant scales for the MMPI-2 items. By using various approaches, they identified 25 new scales (i.e., 23 designated as Specific Problem Scales and 2 as Interest Scales). The 23 Specific Problem Scales were further subdivided into sets targeting somatic/cognitive complaints, internalizing and externalizing, and interpersonal functioning. The two interest areas are Aesthetic-Literary Interests (AES) and Mechanical-Physical Interests (MEC). Finally, the MMPI-2-RF contains revised versions of the Psychopathology Five (PSY-5), which are Aggressiveness-Revised (AGGR-r), Psychoticism-Revised (PSYCH-r), Disconstraint-Revised (DISC-r), Negative Emotionality/Neuroticism-Revised (NEGE-r), and Introversion/Low Positive Emotionality (INTR-r).

As stated, the MMPI-2-RF is not meant as a substitute for the MMPI-2 but, rather, as an additional resource for personality assessment. It can be hand scored, but Ben-Porath and Tellegen (2008) recommended computer scoring and there is software for converting MMPI-2 answers to MMPI-2-RF results. Although Tellegen and Ben-Porath (2008) provided substantial validation evidence concerning the MMPI-2-RF, Butcher (2011) contended that the validation evidence is insufficient. Certainly, the MMPI-2-RF will be the focus of research in the coming decade, and it will be interesting to see what researchers and clinicians say about the MMPI-2-RF.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In conclusion, in order to use the MMPI-2-RF or the MMPI-2, a clinician must be licensed to practice psychology and have training and supervised experience with these types of clinical experience. Most counselors will not have this background, but they may be part of a multidisciplinary team and need a working knowledge of the MMPI-2 or the MMPI-2-RF. The descriptions provided in this book are introductions to these complex and sophisticated instruments.

Other MMPI Related Instruments

Another instrument associated with the MMPI is the *California Psychological Inventory* (CPI). The CPI, which is now in its third edition (Gough & Bradley, 1996), drew almost half of its 434 questions from the MMPI. The CPI scales stress aspects of personality that are more normal as compared with the clinical scales of the MMPI. For example, the CPI provides measures of 20 scales, such as Sociability, Self-Acceptance, Empathy, Well-Being, and Responsibility. It is also possible to obtain scores on five factors of personality: Extroversion, Personal Integrity, Conventionality, Flexibility, and Emotionality. The CPI can be used in many counseling settings with clients 13 years of age and older. The CPI has a rich history and has been researched for more than 40 years. A computer-generated report can be produced for a shortened version of the CPI (260 items) or for the original version (434 items). The CPI does not appear, however, to be as popular as some other instruments, such as the Myers-Briggs Type Indicator.

Another instrument developed from the MMPI is the *Personality Inventory for Children—Second Edition* (PIC-2; Lachar & Gruber, 2001), which is completed by a child's parent or caregiver. Similar to the MMPI-2, this instrument has validity scales that measure Inconsistency, Dissimulation, and Defensiveness. The adjustment scales are Cognitive Impairment, Impulsivity and Distractibility, Psychological Discomfort, Family Dysfunction, Reality Distortion, Somatic Concern, Delinquency, Social Withdrawal, and Social Skill Deficits. In addition, Lachar and Gruber (1995) published the Personality Inventory for Youth (PIY), which is an extension of the PIC-2. The major difference between the two assessments is that the child or adolescent completes the PIY. The Student Behavior Survey, which is completed by a teacher, complements the other two assessments.

LO5 NEO-PI-3

Some exciting advancements in personality research directly pertain to personality assessment, and one of the instruments that reflects this research is the NEO Personality Inventory-3 (NEO-PI-3; McCrae & Costa, 2010). The NEO-PI-3 is a revised version of the NEO Personality Inventory-Revised (NEO-PI-R; Costa & McCrae, 1992) and it builds on the considerable research on the NEO-PI-R. Substantial research indicates that personality can best be described by five factors (Digman, 1990; Goldberg, 1994; McCrae & John, 1992). Researchers have gathered multiple descriptors of personality, and using factor analysis they found that all these different descriptors fall into one of the five factors (e.g., Digman & Takemoto-Chock, 1981; Goldberg, 1992; McCrae & Costa, 1997). The prominence of these five factors has led some researchers to label them the "Big Five." These Big Five factors have traditionally been numbered and labeled as follows: Factor I, *Surgency or Extroversion*; Factor II, *Agreeableness*; Factor III, *Conscientiousness*; Factor IV, *Emotional Stability or Neuroticism*; and Factor V, *Intellect or Openness to Experience*. A mnemonic for remembering the Big Five factors of personality is OCEAN: O for Openness, C for Conscientiousness, E for Extraversion, A for Agreeableness, and N for Neuroticism.

Not only is there substantial research that personality can be described by these five factors but it also appears that the factors may apply across diverse cultures. McCrae and Costa (1997) found, with a particularly large sample, that the American personality factors

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

showed remarkable similarity to the factor structure in German, Portuguese, Chinese, Korean, and Japanese samples. They contended that their findings, in addition to other research, indicate that the structure of personality is universal. These results are encouraging and suggest that instruments based on the Big Five may be appropriate with clients from diverse backgrounds.

Two researchers have been influential in developing instruments that operationalized the five factors. Costa and McCrae (1992) developed the revised NEO-PI-R, which was modified with the newer instrument being the NEO-PI-3 (McCrae & Costa, 2010). There is also an abridged form of the NEO-PI-3 called the NEO Five Factor Inventory-3 (NEO-FFI-3). The NEO-PI-3 has 240 items, whereas the NEO-FFI-3 has 60 items. The NEO-FFI-3 measures the five domains of *Neuroticism* (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). The NEO-PI-3 measures these same 5 domains and 30 traits that are grouped within the domains (i.e., six traits associated with each domain). There are two versions of both the NEO-FFI-3 and the NEO-PI-3: Form S for self-reports and Form R for observer ratings. Caprara, Barbaranelli, and Comfrey (1995) found the NEO to be among the personality tests most heavily researched in the 1990s.

One of the major changes for the NEO-PI-3 as compared with the NEO-PI-R is the newer version can be used with adolescents and adults (i.e., 12 years and older) as compared to only adults. The answer sheets can be hand scored or scanned, and the scores are plotted on a profile.

There is some debate about the appropriate names for the five factors. In the NEO-PI-3, McCrae and Costa (2010) have labeled their domains Neuroticism, Extroversion, Openness, Agreeableness, and Conscientiousness. The *Neuroticism* (N) scale provides a measure of adjustment or emotional stability. High scores are associated with a general tendency to experience negative feelings, such as fear, sadness, embarrassment, anger, guilt, and disgust. Although clinicians make distinctions between types of emotional distress (e.g., anxiety, anger, depression), McCrae and Costa have found that if individuals are prone to one of these emotional states, they are more likely to experience the others. Thus, this domain measures a tendency toward coping poorly with stress, difficulty controlling impulses, and a proclivity toward irrational thoughts.

The *Extroversion* (E) scale concerns the individual's tendency to be sociable, assertive, active, and talkative. High scores are associated with being upbeat, energetic, and optimistic. Extraverts tend to be excitement seeking and to have cheerful dispositions. McCrae and Costa (2010) have found that introverts are not unhappy, hostile, or shy. Their studies indicated that introverts are reserved rather than unfriendly, independent as compared with followers, and even-paced as compared with sluggish. Although they are not likely to exhibit the exuberant high spirits of extraverts, introverts are generally not unhappy or pessimistic.

The third dimension of personality, according to McCrae and Costa (2010), is *Openness to Experience* (O). Elements associated with openness are an active imagination, aesthetic sensitivity, attentiveness to inner feelings, intellectual curiosity, and a preference for variety. Open individuals are interested in both the inner and outer worlds; where low scores are associated with preference for the familiar. This is perhaps the most controversial of the domains on the NEO-PI-3. Other researchers labeled this factor Intellect and contended that the Openness label does not adequately address the factor (Goldberg, 1994; Widiger, 1992). McCrae and Costa argued that this is the most researched scale of the NEO-PI-3 and that a preference for the familiar as compared with a preference for novel experiences is one of the five factors of personality.

The *Agreeableness* (A) scale measures a tendency to be sympathetic, a desire to help others, and a belief that others will reciprocate similarly. Low scores on this scale are associated with a tendency to be antagonistic, egocentric, and skeptical of others' intentions. McCrae and Costa (2010) posit that extremes at either end of this scale are problematic.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Low scores are associated with being disagreeable, egocentric, skeptical, and competitive; high scores are associated with dependency problems.

The last domain is *Conscientiousness* (C), which is related to the individual's ability to control impulses and a will to achieve. This domain measures self-control and the ability to plan, organize, and carry out tasks. Low scores are associated with being more lackadaisical in the pursuit of goals. Extremely high scores can indicate fastidious behaviors, compulsive neatness, and workaholic tendencies.

Each of the five domains (N, E, O, A, C) has six facets. As reflected in Figure 12.2, the facets for the Neuroticism domain are Anxiety, Angry Hostility, Depression,

FIGURE 12.2
NEO-PI-3. Reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, Florida 33549, from the NEO™ Personality Inventory-3 (NEO™-PI-3) by Paul T. Costa Jr., PhD and Robert R. McCrae, PhD, Copyright 1978, 1985, 1989, 1991, 1992, 2010 by Psychological Assessment Resources, Inc. (PAR). Further reproduction is prohibited without permission of PAR

NEO-PI-3 Scale Scores

Norms: Combined Gender, Adult

Factor Scales

Scale	T Score	Percentile
(N) NEUROTICISM	79	99
(E) EXTRAVERSION	31	2
(O) OPENNESS TO EXPERIENCE	37	9
(A) AGREEABLENESS	55	69
(C) CONSCIENTIOUSNESS	45	30

Facet Scales

Neuroticism Facets

Scale	Raw Score	T Score	Percentile
(N1) Anxiety	27	70	97
(N2) Angry Hostility	19	61	86
(N3) Depression	28	77	99
(N4) Self-Consciousness	30	83	99
(N5) Impulsiveness	21	63	90
(N6) Vulnerability	24	80	99

Extraversion Facets

Scale	Raw Score	T Score	Percentile
(E1) Warmth	16	36	8
(E2) Gregariousness	7	30	2
(E3) Assertiveness	7	31	2
(E4) Activity	11	34	5
(E5) Excitement-Seeking	7	30	2
(E6) Positive Emotions	8	25	0

Openness to Experience Facets

Scale	Raw Score	T Score	Percentile
(O1) Fantasy	10	35	6
(O2) Aesthetics	14	46	34
(O3) Feelings	22	54	65
(O4) Actions	9	31	2
(O5) Ideas	8	32	3
(O6) Values	16	40	15

Agreeableness Facets

Scale	Raw Score	T Score	Percentile
(A1) Trust	21	53	61
(A2) Straightforwardness	16	41	18
(A3) Altruism	20	41	18
(A4) Compliance	16	49	46
(A5) Modesty	28	70	97
(A6) Tender-Mindedness	17	41	18

Conscientiousness Facets

Scale	Raw Score	T Score	Percentile
(C1) Competence	15	33	4
(C2) Order	15	41	18
(C3) Dutifulness	17	37	9
(C4) Achievement Striving	16	42	21
(C5) Self-Discipline	16	40	15
(C6) Deliberation	19	52	57

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Self-Consciousness, Impulsiveness, and Vulnerability. There are eight items for each of the facet scores. Three statements at the end of the self-report instrument address the validity of the profile. These, however, are simplistic statements: asking the individual (a) if he or she responded honestly and accurately, (b) if he or she responded to all statements, and (c) if he or she entered responses in the correct areas.

Although unrelated to the NEO-PI-3, there is an instrument for children that measures the five factors of personality. The FFPI-C: Five Factors Personality Inventory-Children (McGhee, Ehrler, & Buckhalt, 2007) is for children and adolescents (9-0 to 18-11). It assesses the traits of Agreeableness, Extraversion, Openness to Experience, Conscientiousness, and Emotional Regulation.

The reliability of the five domains of the NEO-PI-3 is supported by internal consistency estimates that range from .89 to .93. The internal consistency coefficients for the 30 facets scores are somewhat lower and range from .54 to .83. Concerning the validity of the NEO-PI-3, there is a bibliography available from the publisher's website (www.parinc.com) that contains more than 2,500 publications that have used the NEO inventories. Much of the validation information was gathered using the NEO-PI-R, but only 37 of the 240 items were changed for the NEO-PI-3. There is considerable support for the five factors of personality. In fact, Allik and McCrae (2004) found support for the 5 domains of the NEO-PI-R across 36 different cultures. There is, however, significantly less research conducted with adolescents, and counselors should be cautious in using the NEO-PI-3 with this population.

In assessing the Big Five Factors of personality, a counselor should be aware of research regarding the stability of personality across the lifespan. In a meta-analysis of longitudinal study of the Big Five factors of personality, Roberts, Walton, and Viechtbauer (2006) found that conscientiousness, emotional stability, and social dominance (a facet of extroversion) increased during early adulthood (i.e., ages 20 to 40). Furthermore, they found personality changes in older adults, such as increases in agreeableness throughout adulthood including adults in their 70s. The findings of Roberts et al. (2006) contradict the myths that personality change occurs most dramatically in adolescence and that personality is to a large degree set by the age of 30. It should be noted, however, that their findings concern mean differences in personality and that each individual may not necessarily show these changes in conscientiousness, emotional stability, and social dominance in early adulthood.

In the manual, McCrae and Costa (2010) discussed a number of usages of the NEO-PI-3 such as understanding the client, assisting in empathy and rapport building, providing feedback and insight, and selecting appropriate treatment. Although the NEO-PI-3 was not designed for assessing psychopathology, McCrae and Costa asserted that the NEO-PI-3 can be relevant to certain diagnoses because of its focus on personality traits. McCrae and Costa contended that personality traits are most obviously relevant to personality disorders. Clark (2007) contended that the method used for diagnosing personality disorders (i.e., 10 categorical disorders) in the *Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition, Text Revision* (American Psychiatric Association, 2000) is not supported by research and suggested that the facets on the NEO-PI-3 may be an alternative in terms of deepening our understanding of the traits associated with personality disorders. There is currently some debate about the best manner in which to assess and diagnose personality disorders, and, therefore, it is difficult at this point to determine the role of NEO-PI-3 in that process. Nevertheless, the NEO-PI-3 builds on a solid measure of normal personality and has applicability in many counseling settings.

LO6 Myers-Briggs Type Indicator

The Myers-Briggs Type Indicator® inventory, authored by Isabel Briggs Myers and Katharine C. Briggs (MBTI®), is a widely used instrument based on Jungian theory. The theory suggests that variations in behavior are related to basic differences in the ways individuals prefer

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

to perceive and then make judgments about what they have perceived. Furthermore, Carl Jung proposed that individuals have different attitudes (i.e., introversion or extroversion) in which they use their perceptions and judgments. The different forms of the MBTI inventory are designed to assist people in understanding their preferences as measured on four dichotomies (see Figure 12.3). There is the Myers-Briggs Type Indicator Step I (Form M), which is the most commonly used instrument and will be referred to in this book as the MBTI. There is also the Myers-Briggs Type Indicator Step II (Form Q), which will be referred to as the Step II. Although not directly related to either form of the Myers-Briggs Type Indicator, there is a children's instrument, the Murphy-Meisgeier Type Indicator for Children (Murphy & Meisgeier, 1987), which provides similar results in terms of typology. The Murphy-Meisgeier Type Indicator for Children contains 70 items and is for children in second through eighth grades.

FIGURE 12.3
Myers-Briggs Type Indicator Report Form M. Modified and reproduced by special permission of the Publisher, CPP, Inc., Mountain View, CA 94043 from MBTI® Report Form M by Peter B. Myers and Katharine D. Myers. Copyright 1943, 1944, 1957, 1962, 1976, 1977, 1983, 1984, 1987, 1991, 1993, 1998, 2011 by Peter B. Myers and Katharine D. Myers. All rights reserved. Further reproduction is prohibited without the Publisher's written consent

mbti

STEP I / REPORT FORM FORM M

Name _____ Date _____

The MBTI® instrument reports your preferences on four dichotomies. There are two opposite preferences on each dichotomy, as shown below.

E–I Dichotomy Where you like to focus your attention	**E Extraversion** You prefer to focus on the outer world of people and things	**I Introversion** You prefer to focus on the inner world of ideas and impressions
S–N Dichotomy The way you like to look at things	**S Sensing** You tend to focus on the present and on concrete information gained from your senses	**N Intuition** You tend to focus on the future, with a view toward patterns and possibilities
T–F Dichotomy The way you like to go about deciding things	**T Thinking** You tend to base your decisions primarily on logic and on objective analysis of cause and effect	**F Feeling** You tend to base your decisions primarily on values and on subjective evaluation of person-centered concerns
J–P Dichotomy How you deal with the outer world	**J Judging** You like a planned and organized approach to life and prefer to have things settled	**P Perceiving** You like a flexible and spontaneous approach to life and prefer to keep your options open

YOUR REPORTED TYPE AND PREFERENCE CLARITY CATEGORY

Your reported type comprises four letters representing the four preferences you chose. Your preference clarity category (pcc) shows how consistently you chose one preference over the other. High points indicate a clear preference; note, however, that the pcc does not measure abilities or development. To determine your pcc, follow these steps:

1. Refer to the "Points" chart on page two of the answer sheet. For each dichotomy, identify the preference with the greater number of points, and record that letter and number in the "Your Reported Type" column.
2. For each dichotomy, circle in the chart below the range that includes the number next to your preference.
3. Identify the preference clarity category ("slight," "moderate," etc.) shown above each circled range and record it below. If you did not answer all of the items, your points may be lower than the lowest range of numbers on the chart. If so, use "slight" as your pcc.

YOUR REPORTED TYPE	DICHOTOMY	PREFERENCE CLARITY CATEGORY *Raw Points Ranges* Slight	Moderate	Clear	Very Clear	YOUR PREFERENCE CLARITY CATEGORY
	E–I	11–13	14–16	17–19	20–21	
	S–N	13–15	16–20	21–24	25–26	
	T–F	12–14	15–18	19–22	23–24	
	J–P	11–13	14–16	17–20	21–22	

Each type, or combination of preferences, tends to be characterized by its own interests, values, and unique gifts. On the back of this page is a brief description of each of the sixteen types. Find your reported type and see whether the description fits you. If not, the person who administered the MBTI® instrument to you can help you identify a better-fitting type. Whatever your preferences, you may still use some behaviors that are characteristic of contrasting preferences. For a more complete discussion of the sixteen types and applications, see *Introduction to Type®*, 6th ed. (Myers, I. B., 1998, CPP, Inc.) or *Gifts Differing* (Myers, I. B., with Myers, P. B., 1995, Davies-Black® Publishing).

CPP 3803 East Bayshore Road • Palo Alto, California 94303 • 800-624-1765 • www.cpp.com 6134

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The MBTI is a typology instrument in which scores on the four dichotomies result in individuals being categorized into 16 psychological types. The manual indicates the MBTI is appropriate for individuals 14 and older who can read at the eighth-grade level (Myers, McCaulley, Quenk, & Hammer, 1998) the Myers-Briggs Type Indicator Step I (Form M) is a comparatively short personality assessment with only 93 items, which is one of the reasons it is used frequently in organizational consulting. The client's answers to these 93 items result in scores on four dimensions that are then interpreted in conjunction with 16 types (e.g., ENFP, ISTJ).

The first dichotomy, or dimension, on the MBTI indicates a client's preference in attitudes in terms of *extroversion-introversion*. Jung described extroversion and introversion as being complementary attitudes. He suggested that everyone feels a pull between extroversion and introversion; thus, the MBTI is designed to measure an individual's preference along a continuum. The extroverted attitude (E) on the MBTI is not a simple indicator of being socially outgoing. Rather, the extroverted attitude reflects that an individual's attention seems to flow out or be drawn out to people and objects in the environment. An introverted attitude (I) is a preference for drawing energy from the environment and consolidating it within one's own position. Extroverts prefer to focus their energy on the outer world, whereas introverts prefer to focus their energy on the inner world of ideas and concepts.

The client's preference in terms of perceiving information is indicated on the second dimension, which is labeled the *sensing-intuition* dimension. A higher sensing (S) score indicates a preference for perceiving using the five senses. These individuals use the senses to perceive what exists, and they often develop skills in perceiving detail, observing the situation, and remembering specifics. Those with higher scores on the intuition (N) side of the continuum perceive more by using "hunches" and perceptions of patterns. Those with an intuition preference perceive relationships, meanings, and possibilities. As an example, a sensing person will be able to describe the color of the carpet and the amount of wear on the furniture, whereas the intuitive person will contend that the room had a depressed feel to it.

The third dichotomy on the MBTI concerns how the individual processes or makes judgments about the information brought in through the second dimension. This third dimension measures where individuals prefer to make judgments on the *thinking-feeling* continuum. The thinking (T) function is a preference for processing perceptions using rational, cognitive abilities. A higher T score reflects a preference toward making judgments logically and impersonally. A preference toward feeling (F) indicates that an individual primarily uses personal or social values in making judgments. Individuals with higher F scores are attuned to the values of others, as well as their own, when making decisions.

The last dimension describes the process an individual primarily uses in dealing with the outer world. The MBTI assessment focuses on how individuals perceive information and how they make judgments about that information. Thus, this dimension reflects the *judging-perceiving* preference in dealing with the external world. Those who prefer judging (J) favor arriving at a decision or conclusion (using either thinking or feeling) more quickly than those who prefer perceiving (P). Those who prefer perceiving like to perceive situations and, thus, will favor putting judgments off to spend more time perceiving. The judging-perceiving dimension was not explicitly addressed in Jung's discussions of psychological types, and there are questions about whether it is truly independent from the other dimensions (Hood & Johnson, 2007).

The preferences on the four dimensions of the MBTI assessment result in a four-letter code. The different methods of combining the letters produce 16 personality types. For example, an ESFP reflects an extrovert who uses senses to perceive information and feelings to process the information and whose orientation to the outer world is perception.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

An INTJ is an introvert who perceives using intuition and makes judgments by thinking and whose orientation to the outer world is judgment. Counselors, however, need to be careful not to pigeonhole clients based on their personality types, because there is wide variation within each of the 16 personality types. There can often be an inherent link between types and stereotypes, and counselors using any kind of typology instrument need also to consider the uniqueness of the individual.

One of the reasons for the popularity of the MBTI inventory rests in its philosophical stance that all types are valuable and that the preferences are not correct or incorrect. The developers of the MBTI stressed that the preferences are simply differences rather than indicators of health or pathology. Myers explained this concept in detail in *Gifts Differing* (Myers & Myers, 1980), which the MBTI manual suggests as a basic reference for anyone using the instrument. A variety of other resources are also available to aid in the interpretation of MBTI results.

On most personality tests, there are concerns about *social desirability*, or the tendency for individuals to answer questions in a manner that is socially acceptable rather than in a manner that truly reflects their personality. The developers of the MBTI attempted to control for this by weighting items differently depending on measures of the popularity of responses. Item response theory (IRT) has had a significant impact on achievement testing. It also appears to be having a similar impact on personality assessments, as evidenced by the fact that the 1998 revision of the MBTI inventory incorporated IRT. In this revision, the instrument's developers examined the item characteristic curves (ICC) for each item (refer to Chapter 4).

There is detailed reliability information on the MBTI that should be analyzed in depth by a clinician using it. The split-half reliability coefficients and the measures of internal consistency are mostly in the .90s, with some coefficients in the .80s. These coefficients, however, are based on continuous scoring of the dimensions rather than on the dichotomous types of scoring. Both Fleenor (2001) and Mastrangelo (2001) criticized the measures of internal consistency, citing research indicating that after four weeks, an average of only 65% of the sample had the same four-letter code that they had on the first testing. Using 70 reliability coefficients in a reliability generalization study (see Chapter 3), Capraro and Capraro (2002) found mean coefficients ranging from .76 to .84 for each of the four dimensions and an overall instrument average coefficient of .82.

As indicated earlier, there are various forms of the Myers-Briggs Type Indicator, and the most recent iteration is Form Q or Step II. With Step II, each dimension is further delineated by five facet scores. The facet scores are designed to facilitate more detailed knowledge about the individual's personality. For example, the facets of the extroversion-introversion dimension are initiating-receiving, expressive-contained, gregarious-intimate, active-reflective, and enthusiastic-quiet. Step II involves 51 more items than Step I for a total of 144 items. The dimensions and facets scores on Step II were developed using IRT. The items for the facets vary from 5 to 9 items per facet. As Step II is a newer instrument, it involved a new norming group of 1,360 individuals, which Hess (2003) claimed needed to be expanded given the use of 16 types for interpretations. Considering that it is a comparatively new instrument, the validity evidence is also limited (Lanning, 2003). Step II can only be scored online, but the counselor or his or her organization must also subscribe to SkillsOne.com.

In using the MBTI or the MBTI Step II, a counselor needs to consider whether the use of Jungian theory is appropriate for the client and the situation because, unlike some personality instruments, this instrument is based on a theory. The Myers-Briggs Type Indicator assessment has been used in a wide variety of settings and for diverse reasons (e.g., family counseling, career counseling, team building). The widespread use of the MBTI is one of its major problems. As discussed in Chapter 4, an instrument is neither valid nor

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

invalid; rather, validity concerns the support for specific interpretations and the uses of an instrument's results. A counselor who considers using the MBTI or Step II needs to be familiar with the validity evidence and determine if the validation evidence supports how he or she is considering using the instrument.

Other Standardized Personality Instruments

The three personality instruments previously discussed are commonly used standardized assessments in counseling and are examples of the different methods of constructing instruments. The empirical criterion keying method was used to construct the MMPI-2; the NEO-PI-3 was constructed using factor analysis; and the MBTI is an example of an instrument based on personality theory. Another commonly used personality instrument is the *Sixteen Personality Factor Questionnaire* (Cattell, Cattell, & Cattell, 1993), which is commonly called the 16PF. This instrument was constructed using factor analysis, which Cattell started by compiling personality adjectives. Cattell identified 17,953 adjectives that were reduced to 4,504 "real" traits. An examination of commonalties reduced this to 171 traits. Through a series of factor analyses, Cattell first identified 12 pure factors, and then 16 pure factors, of personality. The 16PF is now in its fifth edition and includes measures of 16 factors and 5 Global Factors: Extroversion, Anxiety, Tough-Mindedness, Independence, and Self-Control. Users of the 16PF should be aware that more current technical information and updated norms are available (Maraist & Russell, 2002). The 16PF is an instrument for adults, although a similar instrument, the 16PF Adolescent Personality Questionnaire (Schuerger, 2001), is available for adolescents.

Douglas Jackson influenced the area of personality assessment by constructing the *Jackson Personality Inventory—Revised* (JPI-R; Jackson, 1994), which contains 15 subscales organized in terms of the following five higher-order clusters: Analytical, Emotional, Extroverted, Opportunistic, and Dependable. The item selection procedures for the instrument were technically sophisticated. This instrument is easy to score by hand, and software is now available for generating computer reports. The Jackson instruments are psychometrically sound and well-researched instruments, but they are not used in counseling practice as often as some other instruments.

Limitations of Standardized Personality Inventories

One of the limitations of standardized personality inventories is that the majority of them are self-report instruments. Research has shown that when individuals are motivated to distort the results of a self-report personality inventory, they are quite adept at it (Strickler, 1969; Wiggins, 1966). However, outright lying on a self-report personality measure is not that common in counseling and hiring situations (Schwab & Packard, 1973). Some individuals may be motivated to "fake bad" as part of a trial defense, or some clients may attempt to "put the best foot forward" in order to present a more positive image to the counselor. Bornstein, Rossner, Hill, and Stephanian (1994) found that when a personality trait could be easily identified from the item (i.e., the item had high face validity), then that item was more likely to receive a socially desirable response. Although providing a socially desirable response is a common response set to personality instruments, it is not the only one. Some individuals who have a tendency toward acquiescence commonly select the "true" or "yes" response. Other individuals may have a tendency to select the uncommon or less socially desirable responses. Still others may have a tendency to select the extreme choices on almost all items.

As the previous discussion indicates, some instruments have validity scales that address possible misrepresentation. Another method for controlling for faking and response set use is to inform the client of the inventory's purpose and how the results will be used

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(e.g., for the counselor to understand a client better or for a couple to have a better understanding of their different personalities). In addition, it is often helpful to instruct the client to answer each question honestly. Having clients focus on each question sometimes decreases the tendency to endorse items in a similar manner.

L07 Projective Techniques

As indicated earlier, counselors often use a combination of informal and formal techniques in assessing personality. Within formal personality assessment, there are two major categories: standardized instruments and projective techniques. Projective techniques address some of the limitations of standardized instruments by providing clients with a relatively unstructured stimulus to which they respond. The intent of this technique is to make it more difficult for clients to fake their responses because they will be unsure what the unstructured stimuli are designed to measure. With projective techniques, there are no right or wrong answers because there are no specific questions. The client responds to the stimuli, and the examiner examines those responses and interprets their meanings.

Many individuals trace the use of projective assessments to the early 1900s and the rise of psychoanalytic theory. The psychoanalytic concept of projection concerns individuals' tendency to project their own drives, defenses, desires, and conflicts onto external situations and stimuli. Thus, projective techniques encourage clients to project their personality through their responses to the unstructured stimuli. Projective techniques are thought to uncover more of the client's unconscious and, thus, provide an indication of covert or latent traits.

Projective techniques originated in clinical settings such as psychiatric hospitals and many remain in these settings. Some projective techniques, such as drawings and sentence completions, are also used in counseling settings. There is significant subjectivity in the interpretation of projective techniques, and extensive training is needed to use these instruments appropriately. In addition, there is no professional consensus on how any of these techniques should be specifically interpreted. Lindzey's (1959) categorization of projective techniques into five categories still provides a useful framework. The five categories of projective techniques are *associations* (inkblots, word associations), *construction* (of stories or biographies), *completions* (of sentences or cartoons), *arrangement or selection* (toys or verbal options), and *expression* (drawing or dance).

Association Techniques

Rorschach Inkblot Test. The Rorschach Inkblot Test is a well-known association technique. The test consists of ten inkblots, with five cards in shades of black and gray; two in black and gray with additional touches of bright red; and the remaining three in several pastel colors. There are numerous methods or systems for using these 10 cards, but typically the inkblots are presented in two phases. In the first phase, the association phase, the cards are presented and the examinee responds. In the second phase, the inquiry phase, the examiner goes back through the cards and, in a nondirective manner, questions the examinee on what prompted his or her responses. The information gathered in both phases is used in the scoring and interpretation of the responses.

One of the more widely used interpretation systems is Exner's Comprehensive System, which is based on research with the Rorschach Inkblots. In fact, this system is so detailed that some argue that the Rorschach is not a projective instrument when the interpretation is based on it. Exner's system takes into account the location, determinants, and content in terms of scoring and interpreting responses. Location refers to the section of the inkblot that the examinee's response focuses on. Some individuals will focus on the entire blot, whereas others will focus on common details, and still others will describe unusual details,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

white space, or some combination thereof. The *determinants* include the form, movement, color, and shading of the blot. For example, there is some evidence that the type of movement, whether human, animal, or inanimate, is associated with different psychological processes. The last scoring category is *content*, which looks at common popular responses as compared with other less common views concerning the content of the inkblot. Exner and his colleagues gathered considerable psychometric data on this interpretative system. The reliability coefficients of this system are significantly higher than those of other systems. There is also now better validity information than previously existed. No matter what scoring and interpretive system is used, however, the Rorschach Inkblots should never be used without extensive training and supervision. Furthermore, there is significant controversy associated with the Rorschach, with some researchers (Hunsley & Bailey, 1999) claiming, "There is currently no scientific basis for justifying the use of the Rorschach scales in psychological assessment" (p. 266). Others (e.g., Hiller, Rosenthal, Bornstein, Berry, & Brunell-Neuleib, 1999), however, concluded that the Rorschach is equally as valid as the MMPI for most professional uses. It should be noted that Maruish (2004) reported a decline in use of the Rorschach because it is extremely time consuming. Also, most counselors will not have training in the interpretation of the Rorschach and, therefore, should not use it.

Construction Techniques

Thematic Apperception Test. The Thematic Apperception Test (TAT) is an example of a constructive projective technique in which examinees construct a story based on a picture shown to them by the examiner. According to Groth-Marnat (2009a), the TAT is typically used to supplement other psychological tests because the elicited stories usually contain clinically rich information. Murray and his colleagues developed the TAT in 1938 as a measure of Murray's theory of needs and presses. The current version (Murray, 1943) has 31 cards—30 black-and-white pictures plus a blank card. Although the TAT is somewhat dated, Camara, Nathan, and Puente (2000) indicated that it is one of the 10 most frequently used instruments by psychologists, which is why counselors may see psychological reports that include the TAT. Depending on the examinee's age and gender, there are 20 specific cards that can be shown to them (19 pictures and 1 blank). The pictures primarily depict people in ambiguous situations (e.g., a young boy is staring at a violin that is placed on a table in front of him). The examinee then constructs a story based on each picture presented. Rarely does someone construct a story for each of the 20 pictures. The stories are recorded verbatim and then analyzed by the examiner.

Like the Rorschach Inkblot Test, numerous methods for scoring and interpreting the TAT have evolved over the years. Some quantitative scoring systems yield acceptable interrater reliability coefficients. These scoring systems, however, are somewhat time consuming and are not used frequently. Although normative data exist on the TAT, clinicians often rely on clinical experience when interpreting the instrument. Relying on this experience can be problematic, however, because research indicates that conditions such as lack of sleep or hunger can influence performance. Another criticism of the TAT is that the pictures are somewhat out of date. However, some clinicians feel that the TAT provides a good method of entering the client's "world," while also providing a way for the client to open up to the counselor. Nevertheless, clinicians' interpretations of client responses are often quite diverse and can even contradict one another. Therefore, as with other projective techniques, counselors need to be cautious in interpreting the responses and they should seek verification from other sources of information.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Completion Techniques

Rotter Incomplete Sentences Blanks, Second Edition. In completion techniques, examinees are provided an incomplete verbal stimulus that they must complete. For example, with sentence completion exercises, a frequently used projective technique, examinees are provided with incomplete sentences and asked to complete the sentence in their own words. Examples of sentences are "I hope …," "People …," "At home …," and "I get upset when …" Some counselors have developed their own incomplete sentence assessments to address various types of issues and problems, including family conflict, transitional issues (such as going to middle/junior high school), career satisfaction, and self-concept.

A carefully constructed and standardized sentence completion test is the *Rotter Incomplete Sentences Blank, Second Edition* (Rotter, Lah, & Rafferty, 1992). This instrument has three levels: high school, college, and adult. The client completes 40 sentences that are scored for conflict, positive responses, and neutral responses. An overall adjustment score can also be calculated from the responses. Boyle (1995) stated that the Rotter Incomplete Sentences Blank is often used in the early stages of counseling, where it is a useful quick screening device. Practitioners often do not score this instrument but, instead, clinically analyze the responses. The norms have recently been updated, and the manual provides new information on reliability and validity.

Arrangement or Selection Techniques

Assessment involving individuals selecting and arranging objects can avoid the difficulties associated with traditional assessment procedures. This is particularly true with young children as they often have difficulty answering questions or completing a paper-and-pencil exercise. Sandplay, or the use of sand trays, has become of some interest to some counselors in the past 20 years. There is some disagreement about whether the sand tray should be used as a diagnostic or clinical assessment instrument or as a therapeutic tool. The International Society for Sandplay Therapy supports the psychotherapeutic application of the sand tray as compared with its diagnostic uses (Mitchell & Friedman, 1994). According to Mitchell and Friedman, the Sandplay literature indicates that five areas are predominantly considered when interpreting a sand tray: (1) how the sand tray is created, (2) the content of the tray, (3) the developmental perspective of a series of trays, (4) the Sandplay story, and (5) the therapist's feeling response.

Most research on Sandplay and the use of sand trays has been individual case studies, with very few empirical investigations. Because this technique has not been investigated extensively, it does not have the empirical base of other projective techniques. Without more research, its use as an assessment tool is questionable. In particular, Murrie, Martindale, and Epstein (2009) found compelling evidence that there is insufficient evidence to use it as an assessment technique in child sexual abuse evaluations. Another problem with the use of sand trays is the lack of standardization. There is not a universal standard on the size of the tray, the amount of water available, or the number, size, or nature of the miniatures used. This lack of standardization severely hampers the research that can be performed or the ability to generalize from one sand tray to another.

Children's selection of, arrangement of, and play with toys have been used as a projective technique for many years. Because play allows for the natural expression of children's thoughts and affects, it lends itself to interpretation at multiple levels. The research on how to interpret play is not as extensive as that of other projective techniques, such as children's drawings. Schaefer, Gitlin, and Sandgrund (1991) included three of the more promising projective play assessments in their book. Interestingly, two of these projective techniques involve working with the entire family. In one of these techniques, the family engages in a collaborative drawing exercise that encourages participation of both adults and children. Each family member contributes to the drawing, which is used to aid in the understanding of the family's functioning. With the second technique, the family uses puppets for reenacting recent problematic situations.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The third technique involves the use of puppets in diagnostic interviews with children and particularly focuses on analyzing the form and content of the child's play with the puppets.

Expression Techniques

Drawing Techniques. It makes logical sense that the methods people use to express themselves may provide a window into their personalities. The method of self-expression that has received the most study in terms of being a projective technique is art or drawing. Oster and Crone (2004) asserted that many clients find drawing less threatening than other types of assessments. In particular, drawing techniques are often used with younger children to facilitate communication, as the clinician often asks questions as the child is drawing. One of the first formal techniques to incorporate drawing was Machover's (1949) *Draw-a-Person Test* (D-A-P). In the D-A-P, the client is first given a piece of paper and a pencil and told to draw a person. Once the first figure has been drawn, the examiner instructs the individual to draw a person of the opposite sex. After the drawing is completed, the examiner either asks the individual to tell a story about each figure or asks a series of questions concerning age, education, occupation, ambitions, and other pertinent details. Machover contended that individuals tended to project acceptable impulses on the same-sex figure and unacceptable impulses on the opposite-sex figure. In addition, certain characteristics of the drawing were thought to be indicative of certain personality traits. For example, hands behind the back were indicative of introversion; dark, heavy lines and shading suggested hostility and aggression; small drawings were related to low self-esteem; and disproportionately large heads indicated organic brain damage. However, Machover's interpretation of the D-A-P has been both questioned and criticized (Swenson, 1968).

There are a number of different approaches to interpreting human figure drawings. In general, interpretations examine pencil pressure, stroke and line quality, lack of detail, placement, erasures, shading, and distortions (Handler, 1996). Experts in this field stress that no single sign has only one fixed and absolute meaning. For example, a child drawing a face with a frown does not always indicate the child is depressed. However, observing and engaging the client in conversation about the drawing can often facilitate the gathering of information through use of spontaneous imagery. It is also important to take into consideration the child's educational background and developmental level.

One of the outgrowths of the human figure drawings is the popular *House-Tree-Person* (H-T-P), which was originally published by Buck (1948), with the latest revision by Buck (1992) and revised by Warren. In H-T-P, the individual draws three separate drawings—one of a house, one of a tree, and one of a person. The individual is then given the opportunity to describe, define, and interpret each drawing. In general, Buck suggested that the house reflects the home situation of the client, the tree represents the self in relationship to the environment, and the person represents the self.

Burns is another influential name in projective drawing techniques. Burns (1987) adapted H-T-P into *Kinetic-House-Tree-Person*, in which the house, tree, and person in action are drawn on one sheet of paper. This technique incorporates the interconnections among the three figures and the person's action. Another technique that Burns (Burns, 1982; Burns & Kaufman, 1970, 1972) is known for is the *Kinetic Family Drawing* (K-F-D). A counselor using this technique asks the child "to draw the family doing something." The incorporation of action into the drawing, according to Burns, allows for a clearer picture of family dynamics. The child can depict the family as an active, functioning unit, which can provide indications of family interactional patterns. Figure 12.4 is an example of a 10-year-old boy's K-F-D. Burns (1982) suggested four major categories in interpreting a drawing: (1) Actions; (2) Distances, Barriers, and Positions; (3) Physical Characteristics

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

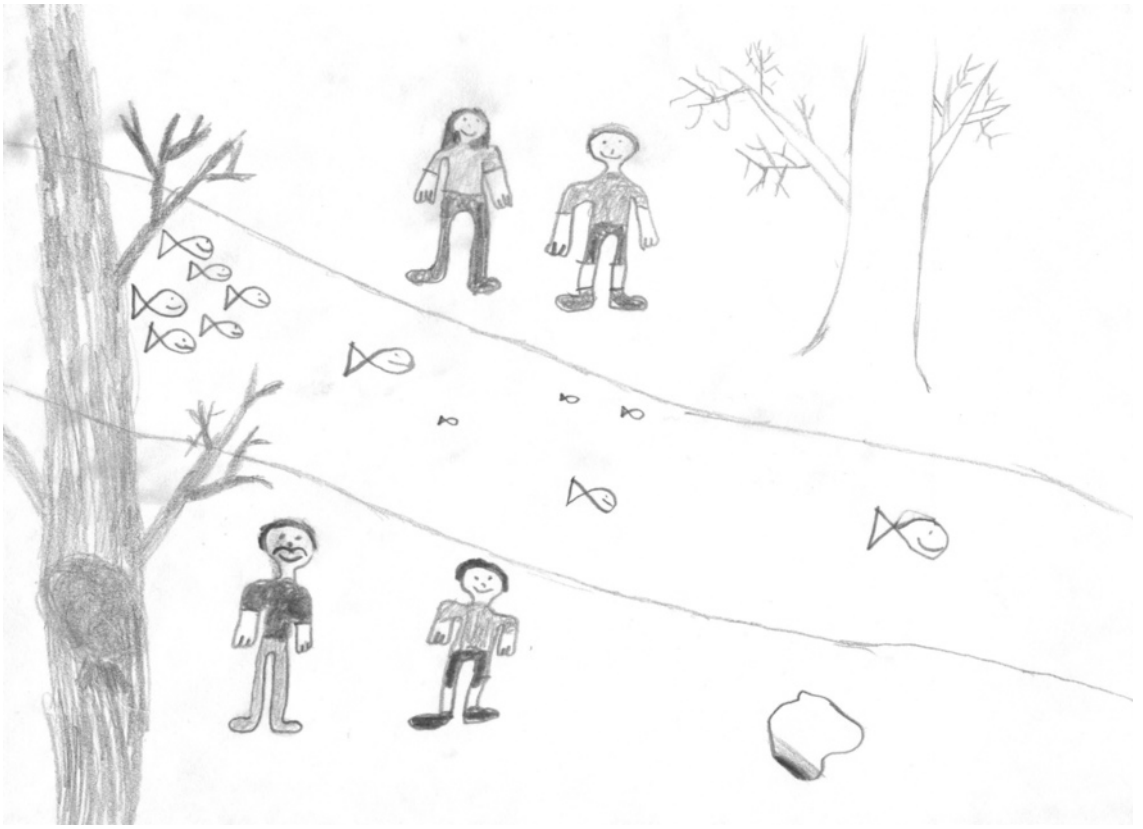


FIGURE 12.4
Example of Kinetic Family Drawing

of the Figures; and (4) Styles. *Actions* concern the overall content, theme, or activity of the picture. *Distances, barriers, and positions* analyze how the family is arranged and what indicators of relationships are shown. *Physical characteristics* examine the formal aspects of the drawing (e.g., expressions, analysis of presentation of body parts, different sizes of the figures). *Style* variables are indicators of emotional problems or disturbance. Interested readers are directed to Burns's publications. A resource for counselors working with children, both in school and nonschool settings, is Knoff and Prout's (1993) *Kinetic Drawing Systems for Family and School: A Handbook*, which explores methods for examining children's difficulties both at home and in school.

Strengths and Limitations of Projective Techniques

Many projective techniques address some of the limitations of standardized personality instruments, in that the former are more difficult to fake. Projective techniques can also sometimes identify complex themes and multidimensional aspects of personality. With some clients, projective techniques can serve as an effective method of establishing rapport. In addition, with young children and nonverbal clients, some of the expressive techniques can provide an opening into the client's world that is difficult with other methods. On the other hand, projective techniques have some significant limitations. The reliability evidence for most of these techniques is quite low, which means there are higher proportions of error in these measures as compared with other measures. Because of these higher proportions of error, a counselor needs to regard the findings more cautiously. Certainly, the examiner and the situation need to be taken into consideration when interpreting the results of a projective technique, because both can have a significant influence on the process. The lack of normative data further compounds the problem. The central concern with the use of any

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

instrument, however, is the validity of that instrument. The validation information on most projective techniques is often meager. In the wrong hands, a projective technique can be dangerous, as some clinicians might see every triangle any client draws as an indication that they were sexually abused. Anastasi and Urbina (1997) suggested that projective techniques should be viewed not as tests but rather as clinical tools. As clinical tools, these techniques inform practice by identifying possible hypotheses to be further investigated in the counseling process. Thus, these tools would not serve to confirm a clinical judgment but rather to generate hypotheses to be further explored and investigated.

LO8 Self-Concept Measures

Some counselors view self-concept as a construct closely related to personality, whereas others define it as a unique aspect of the individual. Self-concept is another area in which there is debate and differing opinions on definition and characteristics. When the topic of self-esteem is added to the discussion, the matter becomes more tumultuous, for self-esteem has become a social/political issue in the past 20 years. Leaving the debate over self-concept and self-esteem to other venues, this discussion addresses some of the prominent measures in this area. In some form, these measures are related to individuals' evaluation of their own performance or feelings about themselves. Counselors use self-concept or self-esteem measures primarily in two ways. First, counselors may want information on client attributes and may use these measures in the beginning of the counseling process to assess the client's level of self-esteem. Second, counselors may want to use these instruments to examine the effect of counseling interventions.

For example, counselors might examine whether a group counseling experience raised the self-esteem level of the participants. It should be noted that research has shown that counseling interventions have not been found to have a consistent effect on self-esteem (Kahne, 1996; Whiston, Tai, Rahardja, & Eder, 2011). This may be because of difficulties associated with measuring self-concept and self-esteem.

Piers-Harris Children Self-Concept Scale, Second Edition

The *Piers-Harris Children Self-Concept Scale, Second Edition* (Piers, Harris, & Herzberg, 2002) is a widely used measure of the self-perceptions of children and adolescents. The scale is composed of 60 items and is appropriate for individuals 7 to 18 years of age. The Piers-Harris provides one total score and six subscale scores that indicate the child's self-esteem in the areas of Physical Appearance and Attributes, Freedom from Anxiety, Intellectual and School Status, Behavioral Adjustment, Happiness and Satisfaction, and Popularity. Kelley (2005) considered the second edition of the Piers-Harris to be an improvement over the original scale and considered it to be one of the best current measures of self-concept. Conversely, Oswald (2005) expressed concern about the overlap of items in many of the different subscales and the essential unidimensionality of the instrument. The lack of a theoretical foundation was also noted by Oswald as a limitation of the *Piers-Harris Children Self-Concept Scale, Second Edition.*

Tennessee Self-Concept Scale

Another frequently used self-concept scales is the *Tennessee Self-Concept Scale—Second Edition* (TSCS-2; Fitts & Warren, 1997). This instrument can be used with children, adolescents, and adults. The results provide two summary scores, with one being a measure of total self-concept and the other a general measure of conflict. There are six specific self-concept scores: Physical, Moral, Personal, Family, Social, and Academic/Work. Furthermore, there are three supplementary scores and validity scores that take into consideration inconsistent responses and faking. Hattie (1998) suggested that the TSCS-2 will continue to be a popular instrument but concluded that additional construct validation evidence would be beneficial.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Summary

Assessing a client's personality is often part of the counseling process because individuals' personalities are frequently intertwined with their problems and issues. In addition, clients' personalities should be considered when making treatment decisions and selecting interventions. Counselors often use both formal and informal personality assessment techniques. Observation is the most common technique in personality assessment. Counselors should attend to the reliability and validity of their observations because personal biases and attitudes can influence their abilities to be skilled observers. Interviewing is another personality assessment technique that counselors frequently use. The quality of the interviewing questions greatly influences the usefulness of the information gathered in an interview. Interviewing suffers from many of the same problems as observation; therefore, counselors need to evaluate the psychometric qualities of their interviews.

Within formal personality assessment, there are two major categories: standardized instruments and projective techniques. The most widely used standardized personality instrument is the MMPI-2. Individuals who are knowledgeable and well trained in the intricacies of this instrument should conduct the interpretation of MMPI-2 results. The MMPI-2 Restructured Form (MMPI-2-RF) is a comparative new addition to the MMPI-based instruments and it will be interesting to see researchers and clinicians reactions to this instrument. Other commonly used personality inventories are the Myers-Briggs Type Indicator, the NEO-PI-3, and the 16PF. These self-report instruments do have some limitations, particularly concerning clients' abilities to manipulate the results and the tendency of clients to use a response set. Projective techniques are more difficult to fake or to be influenced by the clients' response set. Projective techniques, however, usually have lower reliability coefficients than standardized instruments, and many are lacking in sufficient validation information. Furthermore, training is particularly needed in regard to interpreting projective assessments. Self-esteem or self-concept measures are sometimes considered part of personality assessment. However, the area of self-esteem or self-concept assessment has problems in that there is a lack of consensus on what precisely constitutes self-esteem and how to measure it.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 13

Behavioral Assessment

Learning Objectives

- LO1 Explain the premises of behavioral assessment
- LO2 Distinguish between a naturalistic observation and an analogue observation
- LO3 Develop an example of narrative recording
- LO4 Create an example of interval recording
- LO5 Produce an example of event recording
- LO6 Design an example of ratings recording
- LO7 Describe the strengths and limitation of self-monitoring
- LO8 Understand the influences of reactivity
- LO9 Distinguish among the three types of error in behavioral assessment

You are a school counselor at an elementary school, and a teacher approaches you about a student who is exhibiting problematic behavior in class. According to the teacher, the student is too rambunctious. You are unsure what *too rambunctious* means, and the teacher has difficulty describing the behaviors other than noting that the child is continually talking to other students and does not stay in her seat. The teacher wants you to work with the child so that she will concentrate on her school work and spend less time bothering other students. You decide that it would be helpful to observe the student in the classroom so you have a better idea of the problem. You, however, want to ensure that the time you spend observing the child is helpful, and you are interested in different observational techniques that may be useful.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L01 In much of this book, I discuss assessment of career characteristics, personality traits, and abilities. In these types of assessments, the focus is on developing an instrument in which we infer from the results that the client possesses certain characteristics. Behavioral assessment is somewhat different in that the focus is typically on the precise behaviors of the client. It should be noted, however, that in behavioral assessment inferences are sometimes made based on the behaviors exhibited by the client. Nevertheless, the focus of **behavioral assessment** is on observing behavior and often recording that behavior. According to Cohen, Swerdlik, and Sturman (2013), the differences between behavioral assessment and more traditional forms of assessment can be broken down to the *who, what, when, where,* and *how of it.*

Who? The *who* of behavioral assessment is typically focused on the behavior of one individual rather than on assessing a group of individuals to normatively identify some hypothesized trait or state. The client could be a student in an elementary school, a client in a community agency who specializes in substance use, or a couple experiencing marital difficulties. The *who* also includes the assessor, who in behavioral assessment could be the counselor, a teacher, a parent, or even the client. For example, clients are frequently asked to complete behavioral diaries, which are considered a behavioral assessment.

What? Not surprisingly, the *what* that is being assessed in behavioral assessment is one or more behaviors. Often in behavioral assessment there is a **target behavior** that is the focus of the assessment. As will be discussed, it is important that this target behavior be clearly defined, which is sometimes referred to as an *operational definition*. From the example used at the start of this chapter, the counselor would need to clearly identify specific behaviors that are an indication of being too rambunctious. You may want to start thinking of specific behaviors that could be monitored that are reflective of the teacher's concerns. For example, it is much easier to identify when the student is out of his or her seat than it is to monitor some vague behavior such as being *rambunctious*. Other examples of target behaviors are eye contact for individuals who are attempting to improve their social skills or the number of cigarettes smoked for someone attempting to quit smoking. For behavioral assessment, the behaviors need to be clearly identifiable and measurable.

One of the ways of categorizing behavioral assessment is whether the behavior is *direct* or *indirect*. Direct means the behavior is being directly observed and recorded as the behavior occurs. The observation of behavior could be completed by the counselor, an observer, or the client. An indirect method of behavior assessment requires the person to reflect on past behaviors. This is often completed by the client in terms of self-report, but it also could be completed by a parent who is reporting on his or her child's previous behavior. In general, direct observations are preferred over indirect as many people have difficulty recalling their behavior or someone else's behavior in a completely objective manner.

According to Watson and Watson (2009), the dimensions of behavior that are measured in behavioral assessment are the frequency or rate of behavior, the duration of a behavior, the relative intensity with which a behavior occurs, and the latency of a behavior. *Frequency* generally refers to the number of times an event occurs. For example, a counselor could count the number of times a child hits another child. In order to use frequency measures effectively, the target behavior must have a discrete beginning and end so that the behavior can be counted. In addition, the target behavior should not occur too frequently as it is difficult to accurately record behaviors that occur very rapidly. *Rate* typically refers to the number of times per minute that an event occurs, such as a child hitting another child 10 times in five minutes, which would be two hits per minute. Different units of time (e.g., minutes, hours, days) may be appropriate for various target behaviors. *Duration* concerns

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the length of the behavior, and it can be measured in various ways using different coding systems. An example of duration is a child stayed out of her seat in the classroom for four minutes. *Intensity* is more difficult to operationalize and measure. For example, you probably recognize that tantrums can vary in intensity, but it is more difficult to define what is the difference between a more intense tantrum and an average tantrum. Sometimes there are rather simple measures of intensity, such as having the client rate the behavior in terms of intensity on a scale from 1 to 10, with 1 representing the least intense and 10 reflecting the most intense. *Latency* represents the time between the stimulus and the behavior. An example of this would be the time it takes for a student to pick up and open the book from the time the teacher instructed the students to start reading. Latency can be an important behavioral dimension, but it is less frequently the focus of attention in behavioral assessment (Watson & Watson, 2009).

When? Often *when* behavioral assessments are made is when the problematic behavior is most likely to occur. For example, if a student has a tendency to get into fights or physical altercations during recess, then recess is when the behavioral assessment is made. Sometimes, however, it is not possible to identify when the problematic behavior will occur; then the counselor wants to select a time or times when the counselor can get the best sample of behavior. With behavioral assessment, we are usually talking about a sample of behavior, which is limited to a specific time. For example, the school counselor in the earlier vignette might observe the rambunctious student for an hour to get a sample of her behavior. There are different methods of timing for the behavioral assessment, and some of the advantages and disadvantages of these different methods will be discussed later in this chapter. For example, you might simply determine whether a behavior occurs in a certain time period or you might observe a client and count the number of times a behavior occurs during a specific length of time.

L02 **Where?** In a **naturalistic observation**, the counselor observes the client in a typical environment and does not manipulate any aspect of the environment during the observation (Watson & Flamez, 2015). The goal of a naturalistic observation is to observe the person in his or her natural setting, which is intended to get a better picture of how the client typically behaves. Sometimes naturalistic observation is called ecological observation, which emphasizes the interaction between the individual and his or her environment. Examples of naturalistic observations are school counselors observing students in a classroom or a mental health counselor making a home visit to observe a family. Furthermore, naturalistic observation can be either overt or covert. In an *overt* observation, the person being observed knows that he or she is being observed. For example, a counselor who is counseling a family in their home (i.e., home-based services), the family knows they are being observed by the counselor. A *covert* observation is where the client does not realize that he or she is being observed. An example of a covert observation is a school counselor observing a student during recess where the student is not aware that he or she is being observed. A covert observation may provide a better indicator of typical behavior, but it is more difficult to implement in many settings. The naturalistic observation can also be structured or unstructured. Many of the approaches to behavioral observations discussed in this chapter are more structured, but counselors may decide that an unstructured approach to behavioral assessment can provide useful client information.

An analogue observation is another alternative that counselors can consider. In an **analogue observation**, the counselor creates a simulated environment that is reflective of the client's natural environment (Watson & Flamez, 2015). Doing a role-play with a client and closely observing his or her behavior is in essence an analogue observation. The goal of an analogue observation is to create an environment in which the client behaves naturally. In an analogue observation it is more difficult to know if the client is behaving naturally so the counselor needs to be more cautious in interpreting the behaviors observed

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

in an analogue observation. On the other hand, analogue observations give the counselor greater control of the situation as the counselor can design the situation to hopefully maximize the possibilities of important behaviors occurring. For example, the counselor could have a couple discuss a contentious topic in session and observe the couple's behaviors. This analogue observation may provide important insight into how the couple communicates and possibly how they argue.

Why? Conducting a behavioral assessment can provide counselors with good information that can be used in screening, in treatment planning, during treatment, and in evaluation of the counseling services provided. For example, a behavioral assessment, such as having a client note each day how many drinks he has can be used for screening purposes. This same log of drinking could be used for treatment planning as the content may identify some triggers of drinking behavior. According to Watson and Watson (2009), one of the basic premises of behavioral assessment is that it should be linked to intervention. In targeting specific behaviors, the counselors can also use this information during treatment with behavior modification types of interventions. Finally, behavioral assessments can be used to establish behavioral baseline data in which the counselor compares to other behavioral assessments completed during treatment or at the end of treatment. For example, the drinking log done at the start of treatment can be compared to a drinking log completed at the end of treatment. Cohen et al. (2013) indicated that in this time of managed care and frugal third-party payers, insurance companies often prefer behavioral assessments over other assessment strategies.

How? How behavioral assessments are completed will vary depending on the client, the counselor, and the purpose of the assessment. The type of behavioral observation will also vary depending on whether the counselor is more interested in frequency, rate, duration, intensity, or latency. Sattler and Riddell (2014) classify observational methods or techniques into the categories of narrative recordings, interval recordings, event recordings, and ratings recording.

Observational Recording Methods

L03 Narrative Recording

Narrative recordings are sometimes referred to as a running record of behavior or an anecdotal record of behavior. These types of recording typically do not involve any quantitative recording procedures. Narrative recordings can be completed by the counselor, parents, other family members, teachers, or the client. The focus is on recording a description of the behavior that seems noteworthy or important to the observer. For example, a counselor might have a parent record behaviors of a child when the child is acting out so that the counselor has a better idea of the specific behaviors.

Depending on the counselor's needs, narrative recordings can be global, semiglobal, or narrow. For example, the parent could record that the child was playing outside or the parent could be narrower and record that the child was skipping rope in the driveway. Narrative recordings are used to provide the counselor with general information and may be a precursor to more detailed behavioral observations. For example, a counselor who suspects possible substance use problems may start by having the client simply record what he or she does in the evening.

Counselors often observe clients, and, therefore, they may be conducting a narrative recording informally. In order to make it more formal, the counselor would focus on the behaviors and accurately record the behaviors. As indicated earlier a narrative recording is like a running record of the behaviors. Table 13.1 provides an example of a narrative recording conducted by a school counselor on the rambunctious child mentioned in the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 13.1
Example of Narrative Recording

NARRATIVE RECORDING FORM

Student Name: Sally Smith Grade Level: 2nd

Teacher: Mrs. Jones Date: March 14, 2016

Time Begin: 9:00 Time End: 9:30

Time	Students' Behavior
9:00	Takes seat
9:03	Kneels on chair
9:04	Gets out of seat and goes to the back of the classroom
9:05	Looks at books in the back of classroom
9:08	Returns to seat at the teacher's request
9:10	Talks to another child
9:12	Gets paper out of desk
9:13	Kneels on chair
9:14	Sits in seat and appears to be attending to the teacher
9:17	Talks to another child
9:19	Stops talking to the other child upon the teacher's request
9:20	Appears to be attending to the teacher
9:23	Is looking through the desk
9:24	Gets crayons out of the desk and starts drawing on paper
9:28	Puts the paper and crayons back in desk
9:29	Appears to be attending to the teacher

earlier vignette. As the table reflects, the focus of narrative recording is on observing the behaviors and recording them. It is important that counselors record behaviors precisely including verbalizations. The focus on narrative recording is to describe rather than to interpret the behavior, although interpretation and inferences about behaviors may occur later.

There are many advantages to narrative recordings. As the methodology requires that behaviors be recorded sequentially, it is possible to examine tendencies and sequential trends. Furthermore, narrative recordings require little equipment and can often provide useful information. The usefulness of the information, however, is dependent on the quality of the inferences made about the behaviors. The intent of narrative recording is to focus on the specific behaviors so that inference drawn can be tied to specific behaviors.

Interval Recording

L04 **Interval recording** focuses on whether specific behavior(s) occur within a certain interval of time. Interval recording is also referred to as *time sampling, interval sampling,* or *interval time sampling*. The term *sampling* conveys that the counselor is just measuring a sample of time in which to determine if a behavior occurs. With interval recording the observational period is divided into brief segments (e.g., 30 seconds), and the counselor or assessor documents whether the target behavior occurs during that

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

time period. Hence, with interval recording the focus is on the absence or presence of the target behavior. Interval recording is more suitable to highly structured behavioral assessments, but it can also be used less formally. For example, a counselor might consciously break a 5-minute portion of a session into 30-second segments and record whether a client reports a negative thought within each 30-second segment. With interval recording, a counselor needs to first clearly define the target behavior(s) being *sampled*. In our example of the rambunctious child, the counselor may define off-task behaviors as being whenever the child is out of her seat, which is a clearly identifiable behavior. The counselor then needs to select the time of the interval and the approach to interval recording.

Partial-interval time sampling is when the behavior is scored if it occurs at any point during the time interval. For example, if a school counselor is observing the rambunctious child, the counselor could break the 30 minutes into 5-minute intervals and then note if the child is out of her seat during each of the 5-minute intervals. Therefore, the recording is just whether the behavior occurs during the time interval. Table 13.2 provides an example of a partial-interval time sampling approach in which the person observing simply marks an X if the behavior occurred during the time interval and an O if the behavior did not occur during the five-minute intervals. *Whole-interval time sampling* is when the behavior has to last through the *entire* time interval. Using our rambunctious child as an example, the school counselor may observe the student in a classroom for 10 minutes and record each time that the child talks to another child for 30 seconds. If the child talked only for 29 seconds, it would not count in this example of whole-interval time sampling as the child has to talk for the entire 30 seconds in order for the behavior to be recorded. Whole-interval time sampling is appropriate when counselors are interested in both frequency and duration of a specified behavior. At the end of a whole-interval time sampling time, a counselor can tally up the number of intervals where the behavior was evident and determine how much of the observation time was devoted to the behavior.

Interval recording techniques require the counselor to monitor the time as the focus is on whether the behavior occurred during the time interval. With partial-time interval sampling, the behavior needs to occur just once during the interval, whereas, with whole-time interval sampling, the behavior must occur during the entire time interval. The decision on whether to use partial-time interval sampling or whole-time interval sampling rests on two questions. The first concerns what dimension (e.g., rate, duration, latency) the counselor is more interested in. If the counselor is more interested in frequency of behavior, then a partial-time interval sampling approach is probably better. If, however, the counselor is more interested in duration, the whole-time interval

TABLE 13.2
Example of Five-Minute Partial Interval Time Sampling Recording

PARTIAL INTERVAL RECORDING FORM

Student Name: Sally Smith Grade Level: 2nd

Teacher: Mrs. Jones Date: March 14, 2016

Time Begin: 9:00 Time End: 9:30

Behavior observed: out of seat

Time Segment*	1	2	3	4	5	6
Behavior	X	X	O	O	O	O

*Each number represents a five-minute block.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

sampling is probably more appropriate. The second question concerns the guiding purpose of the behavioral observation. When a counselor is interested in whether or not a target behavior occurs, then the more appropriate strategy is probably partial-time interval sampling. When the question concerns what might precede a behavior, then whole-time interval sampling may provide a better indicator as this could also be documented.

Event Recording

L05 One of the methods used in behavioral assessment is **event recording**, where the counselor records the number of times a target behavior or behaviors occur during a specified time period. Like all forms of behavioral assessment, the focus is on behavior. Unlike interval recording, where the focus is on the time, the focus with event recording is on the number of times the behavior occurs. In other words, the counselor waits for the behavior (or event) to occur and then records it. An example of event recording is keeping a tally of each time a client mentions a negative thought in a specified period of time (e.g., 30 minutes). The counselor could then say to the client, "I have noticed in the last 30 minutes, you have mentioned negative thoughts eight times." This sort of behavioral assessment can provide concrete evidence to a client about the frequency of a target behavior. According to Sattler and Riddell (2014), event recording provides a continuous record of behavior(s) and thus is particularly appropriate for measuring discrete responses that have clearly defined beginnings and endings. Instead of trying to record something broad like social skills, with event recording, the counselor would record specific behaviors such as making eye contact, saying hello, or asking someone a question.

The primary piece of data obtained in event recording is frequency of behavior. Table 13.3 provides an example of event recording for a child with social anxiety. This event recording was conducted during recess at an elementary school. As can be seen from Table 13.3, the child does not play with any other children during the recess but does talk to other children twice. As indicated earlier, behavioral assessments can be used for evaluative purposes. Thus, a counselor could observe this same child after treatment and see if this second event recording differed from the first.

Event recording is less appropriate for extremely frequent behavior or for those that vary dramatically in intensity and duration. Furthermore, it does not reflect sequences or temporal patterns unless the recording of events is supplemented. Intensity of behavior can be obtained by dividing the behavior into degrees of intensity. For example, you may break raising one's hand in class to (a) raises hand silently, (b) raises hand and talks in normal voice, (c) raises hand and talks loudly. Latency behaviors can also be recorded, but it requires special equipment (i.e., a stop watch) and a stimulus or cue for the behavior (e.g., the teacher instructs a child to take out a worksheet). Latency measures are useful when the counselor is concerned, for example, about a client's ability to follow instruction. In this case, the counselor would time the child on how long it took him or her to respond to the teacher's instruction.

TABLE 13.3
Example of Event Recording for a Child with Social Anxiety

Behavior	Tally	Total Occurrence
Talks to another child	//	2
Steps away from the teacher	///	3
Cries	//	2
Plays with other children		0

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Ratings Recording

L06 In Chapter 8, I briefly discussed rating scales; however, these rating scales are completed primarily by clients and are typically conducted during the intake process. The **ratings recording** category of behavioral assessment is similar in that rating scales are completed, but often these are completed by either the counselor or some other observer (e.g., parents or teachers). Furthermore, a ratings recording method could be used multiple times during the counseling and not just at the beginning.

Ratings recording tends to be used in evaluating more global aspects of behavior and for quantifying clinical impressions (Sattler & Riddell, 2014). As the examples of rating scales demonstrate in Table 13.4, there is frequently subjectivity involved in ratings recording. Because there is more subjectivity, there are often more problems with interrater reliability. If the raters are trained or if the ratings are made more quantifiable, then interrater reliability may increase. Some of the more formal rating scales, which will be discussed later in this chapter, often use multiple raters (e.g., both parents and teachers) to examine consistencies in ratings. This can also be done with informal ratings recording.

TABLE 13.4
Examples of Rating Scales

Examples of Unipolar Rating Scales

Example 1

How depressed was the client? Place an X on the line that best reflects your rating

Not depressed _____: _____: _____; _____: _____ depressed

Example 2

To what extent was the client aggressive during recess? Circle one.

1	2	3	4	5
not aggressive	slightly aggressive	somewhat aggressive	moderately aggressive	strongly aggressive

Example 3

To what extent did the client appear attentive in class? Circle one

1	2	3	4	5
never	seldom	sometimes	frequently	always

Examples of Bipolar Rating Scales

Example 1

How anxious is the client? Circle one.

1	2	3	4	5
very calm	calm	somewhat calm and somewhat anxious	anxious	very anxious

Example 2

How cooperative is the client? Circle one.

1	2	3	4	5
extremely uncooperative	uncooperative	neither uncooperative nor cooperative	cooperative	extremely cooperative

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The advantages of ratings recording are it is easily completed and does not require the observer to monitor the time, which many of the other techniques require. Ratings recording can also be used over time to see if the ratings change as a result of treatment. The disadvantages are that ratings recording tends to be less reliable than other behavioral assessment techniques. Furthermore, there is not a sequential process to ratings recordings; hence, it is difficult to identify events that may be triggering specific behaviors.

Conclusion

The choice of whether to use narrative recording, interval recording, event recording, or ratings recording will depend on client issues, opportunities to observe behavior, and clinicians' observational skills. Although observing clients in their naturalistic environment has many advantages, it should be noted that behavioral assessments can be easily conducted in sessions. It should also be noted that counselors do not necessarily have to be the observer as others, such as parent or teacher, or even the client, can observe the behavior(s) of interest. Having clients observe their behavior will be discussed more in the next section.

Self-Monitoring

L07 According to Watson and Flamez (2015), **self-monitoring** is defined "as the practice of observing and recording one's own behavior" (p. 312). There are various behaviors that counselors have clients monitor and record, such as alcohol consumption, eating behaviors, negative cognitions, exercise, sleep, bullying, lying, or angry outbursts. In fact, you might consider a behavior that you would like to change and think about how you might monitor and record that behavior. In self-monitoring, behaviors are defined broadly and include actions, thoughts, and feelings. Sattler and Riddell (2014) contended that valuable information can be obtained by even having children conduct a self-monitoring assessment.

In conducting a self-monitoring assessment, the client must first understand the response or behavior being monitored and the reason for the assessment. As mentioned in regard to other behavioral assessment, the target behavior(s) need(s) to be clearly defined. This clarity of definition is particularly important in self-monitoring as the client must be able to discriminate between the presence and absence of the behavior or response. Sometimes clients do not complete self-monitoring assessments, not because they are resistant, but because they are unsure of exactly what they should be monitoring. It is often better to start with a single targeted behavior as compared to various behaviors.

The counselor should also give clear instructions on how to monitor and record the response or behavior. Clients may find it easier to comply with the self-monitoring assessment if they are given a form to complete as compared to having to create some recording procedures. This can be quite simple, such as the form included in Table 13.5. Table 13.5 can be adapted to assess various behaviors, such as alcohol consumption. If Table 13.5 was used for the client to self-monitor alcohol consumption, the client would first put the date, followed by the time he or she began drinking. The client would next fill in the event or what was occurring as the client began drinking. Under response, the client would indicate the number of drinks consumed. In this way, the counselor could get information about whether drinking occurred more frequently during certain days of the week or at certain times of the day. Also by having the client record the number of drinks, the counselor can get a better idea of alcohol consumption. Moreover, the act of recording the drinks may sensitize the client to the amount consumed and trends in drinking times and events.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 13.5
Example of Self-Monitoring Log

Date	Time	Event	Response

It is important with self-monitoring that the client records the behaviors when they are occurring or shortly afterward. Retrospective reporting tends to increase memory errors and other distortions. Sometimes it is also important to have the client record antecedent events as they may provide important information to the counselor. The very act of recording antecedent events and behaviors may result in client insight and understanding.

In working with children, Sattler and Riddell (2014) suggested selecting a positive behavior rather than a negative one if possible. It is also advisable with children to work on only one target behavior. Even small children can record behaviors using techniques such as stickers. It is important, however, that the child clearly understands the specific behavior that merits a sticker. Sometimes electronic devices (e.g., cell phones) can be used with older children and adults to remind them that it is time to self-monitor or to record behaviors.

The practice of self-monitoring is not without its drawbacks and difficulties. There can be problems with reliability particularly if there is confusion about the behavior being recorded. Furthermore, the counselor should consider other client information in considering the validation evidence of the self-monitoring—that is, whether there is other information that either confirms or disputes the information gleaned from the self-monitoring. The counselor should also consider whether the client is underreporting or exaggerating the prevalence of the behavior.

In conclusion, self-monitoring can provide rich information to a counselor if clients are motivated to observe and record their behaviors. Sometimes a counselor will instruct a client to start self-monitoring before the client clearly understands the reason for the exercise. For example, a counselor might have a client record negative cognitions before the client understands the connection between his or her negative cognitions and depression. If the client does not understand the rationale for recording the behaviors, then he or she is less likely to comply. Furthermore, the counselor should consider the client's motivational level. If a client is ambivalent about coming to counseling, then assigning the client a complicated and time-consuming self-monitoring assessment may be unwise.

Reactivity

LO8 According to Cohen et al. (2013), "any discussion of behavioral assessment, and particularly self-monitoring, would be incomplete without a mention of the psychometric issue of reactivity" (p. 479). **Reactivity** concerns the possible changes that may occur in clients' behaviors, thoughts, or performance that are a result of being observed, assessed, or evaluated. For example, if you are having a client monitor his or her food intake, then the client may be more inclined to forego the popcorn at the movie theater. Whereas this example is positive, reactivity can also impact behaviors negatively. For example, if you are having a client record negative cognitions, the client may dwell more on those negative cognitions.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Another example of reactivity is when a client becomes nervous while being observed and the client does not do well on a memory task. Sometimes adequate preparation can counter the effects of reactivity for clients, particularly if the client understands the purpose of the observation. There is also the impact of the counseling relationship and the degree of trust the client has in the counselor.

Formal Behavioral Assessments

Thus far in this chapter, we have mainly discussed counselor-created behavioral assessments, but there are also formal behavioral assessments. In my opinion, counselors are more likely to use informal behavioral assessment, but they also should be aware of formal assessments that have been evaluated psychometrically.

Behavior Assessment System for Children, Second Edition

A widely used rating scale of emotions and behaviors is the *Behavior Assessment System for Children—Second Edition* (BASC-2; Reynolds & Kamphaus, 2004). The BASC-2 focuses on both adaptive and maladaptive behaviors and is sometime used in diagnosis, particularly by school psychologists. The BASC-2 is a multimethod and multidimensional assessment, and clinicians need training and supervision in its interpretation. There are five versions of BASC-2: a Teacher Rating Scale (TRS), a Parent Rating Scale (PRS), a student Self-Report of Personality (SRP), a Structured Developmental History (SDH), and a Student Observation System (SOS) for observing classroom behavior. Three separate levels of rating forms are available for the TRS and PRS (preschool, child, and adolescent) and the SRP (child, adolescent, and college).

With the Teacher Rating Scale and the Parent Rating Scale, the participants rate the child on a four-point scale from "never" to "almost always." The Self-Report of Personality includes both true/false items and ratings from "never" to "almost always." There are validity scales that are similar to the MMPI-2 in that they indicate the validity of the responses and whether someone does not understand the items, is not following the directions, has unusual responses, or is being untruthful. The Behavioral Symptoms Index (BSI) is a broad composite score for the TRS and PRS and is designed to assess the overall level of problem behaviors. The TRS has three composite scores of Externalizing Problems, Internalizing Problems, and School Problems, whereas the PRS has the first two composite scores. The clinical and adaptive scales for the TRS and the PRS depend on the age of the individual being assessed, but they include scales on Aggression, Anxiety, Attention Problems, Atypicality, Depression, Social Skills, and Withdrawal. Similar to the BSI on the TRS and PRS, the Emotional Symptoms Index (ESI) is an overall composite score for the SRP. There are four composite scores for the PRS, which are Inattention/hyperactivity, Internalizing Problems, Personal Adjustment, and School Problems, and there are also scales that depend on the age of the examinee (e.g., Alcohol Abuse).

The SDH is a structured broad-based developmental history instrument that is designed to gather information regarding the individual's social, psychological, developmental, educational, and medical history. The Student Observation System (SOS) is a three-part observational tool. In her review of the BASC-2, Stein (2007) was very positive about the TRS and the PRS and to a lesser extent to the PRS; however, she was critical of the SDH and the SOS.

Achenbach System of Empirically Based Assessment

A widely used set of behavioral assessments is the *Achenbach System of Empirically Based Assessment*, which is a series of instruments that varies from age 1½ through 90+. The most commonly used is the *Child Behavior Checklist System/6-18*, which I will spend the most

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

time discussing. There is also the *Child Behavior Checklist/1½-5*, which is for children who are 18 months old to 5 years old. The CBCL/1½-5 is completed by parents although there is also a version for caretakers or teachers. Both forms have Internalizing, Externalizing, and Total Problem scales. In addition a Stress Problems scale was added in 2010. This scale is designed to discriminate between children who have experienced severe stress as compared to those who have not. Another feature of the CBCL/1½-5 is the Language Development Survey, which, based on parent report, measures possible delays in language acquisition.

The *Child Behavior Checklist System/6-18* (CBCL/6-18) is appropriate for children ages 6 to 18 and is completed by the parents or guardians. The version completed by the teacher is the Teacher Report Form/6-18, and the version completed by the child or adolescent is the Youth Self-Report/11-18 (YSR). The Youth Self-Report is only appropriate for children 11 through 18, and it is at the fifth-grade reading level. According to Watson and Flamez (2015), the CBCL/6-18 is one of the most widely used standardized measure for assessing both maladaptive behaviors and emotional problems. The CBCL/6–18 is a four-page assessment, with the first two pages containing descriptive information. The portion that is scored is on pages 3 and 4, and it consists of 113 items. The CBCL/6-18 boys and girls are scored separately, and there are *T* scores and percentiles for Total Competence, three Competence Scales (Activities, Social, and School), eight Syndrome Scales, and six DSM-Oriented Scales. The Syndrome Scales are Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior. There are also three broad scores related to Internalizing, Externalizing, and Total Problems. It is also possible to get scale scores that are DSM-oriented that have been updated to correspond to the *DSM-5.* Although the DSM-oriented scores may assist in diagnosis, the counselor is warned not to rely on these scales exclusively. Flanagan (2005) complimented the developers of these scales on the refinement of these instruments from an earlier version and was particularly positive about the authors' efforts to obtain a nationally representative sample. The CBCL/6-18 can be hand scored, but it is recommended that counselor use either the PC or web-based software. One of the advantages of the computer-scored assessment is that cross-informant comparison can be made.

For children and adolescents, there is also the Semistructured Clinical Interview of Children & Adolescents (SCICA). The SCICA asks open-ended questions that cover children's activities and school performance, peer relations, family relations, self-perceptions, emotions, and parent/teacher-reported problems. The interview is scored and provides syndrome scales, DSM-oriented scales, internalizing, externalizing, and total problems scores.

With the Achenbach System of Empirically Based Assessment, there are also instruments for adults; although these are used less often by counselors than the CBCL/6-18. The Adult Self-Report (ASR) and Adult Behavior Checklist (ABCL) are for individuals 18 through 59. The Adult Behavior Checklist (ABCL) is for someone who knows the client well, such as a spouse or a therapist. There are new multicultural norms for individuals from many societies. There are also the Older Adult Self-Report and the parallel form, the Older Adult Behavior Checklist. These forms are for adults 60 to 90+. For all of the adult instruments, there are adaptive functioning and strengths scales, syndrome scales, and DSM-5-oriented scales.

Error in Behavioral Assessment

L09 As we have discussed before, no assessment in counseling is perfect and there is usually some degree of error within an assessment. There are many sources of error that can occur in behavioral assessment. For example, the client may not be feeling well and may not be behaving typically. Although training of observers is highly recommended before they

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

conduct a behavioral assessment, errors can still occur. Of the systematic errors that can occur in behavioral assessment, there are three that are worth noting: observer bias, observer drift, and observer fatigue (Watson & Flamez, 2015).

Observer bias occurs when the counselor responds to some extraneous factor in the behavioral observation. This can occur because the target behavior is not clearly defined and the counselor's biases play a role in the observation. For example, if I was asked to observe a child for disruptive behavior, I might consider chewing gum (particularly if he smacks his lips) to be disruptive. Now many people are not bothered by gum chewing and would not consider it to be a disruptive behavior. Hence, my aversion to gum chewing is probably an example of observer bias. Another form of observer bias could be related to gender where disruptive behaviors for girls may be different from disruptive behaviors for boys. Observer bias can be minimized by clear operational definition of the target behavior and training of the observer.

Observer drift is another form of systematic error in behavioral assessment. Observer drift occurs during the observational process, the criteria by which the observer uses to categorize or observe changes. This can also occur more easily when there is more ambiguity in the definition of the target behavior. For example, I might start recording gum chewing as a disruptive behavior but decide after some time that it is really no big deal and that I should not record it. This is an example of observer drift. Sometimes observer drift can be less dramatic where the observations subtly change as the observation progresses.

The third type of systematic error in behavioral observation is *observer fatigue*. This is a common occurrence particularly if target behaviors are infrequent and the observation extends over a long period of time. Observer fatigue can cause counselors to miss target behaviors. You may have experienced this yourself when trying to observe someone for a long period of time (e.g., a long boring class). Counselors can minimize observer fatigue by possibly shortening the length of time of one observation and completing multiple observations, which should give you a better sample of behavior.

Summary

Counselors are constantly observing clients, and the intent of behavioral assessment is to make this observation more systematic and, thus, less prone to error. The focus of behavioral assessment is more on observing the behavior and often recording it. Behavioral assessments are often conducted by counselors, but they can also be conducted by parents, teachers, or the clients themselves. Often in behavioral assessment there is a target behavior that is the focus. This target behavior needs to be clearly defined so the observer can readily identify it. Frequently the dimensions of behavior that are measured are frequency, rate, duration, intensity, and latency. Sometimes the location of the behavioral assessment is in a naturalistic setting, while in other situations it is an analogue observation.

There are various methods of recording behavioral assessments, and this chapter covered narrative recording, interval recording, event recording, and ratings recording. Narrative recordings are sometimes referred to as running records of behavior or as anecdotal records of behavior. With interval recordings the focus is on an interval of time and whether a specific behavior occurs. The primary piece of data obtained in event recording is frequency of behavior. Ratings recording requires the use of rating scales to document the behavior.

Counselor can have clients observe and record their own behavior, which is frequently referred to as self-monitoring. It is important in self-monitoring that the client has a clear

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

definition of the behavior(s) being monitored and receives specific instructions on how to record the behavior. In all types of behavioral assessment, but particularly with self-monitoring, reactivity is an issue. Reactivity is related to the possible change that may occur in clients' behaviors, thoughts, or performance that are a result of being observed, assessed, or evaluated.

This chapter includes an overview of two commonly used formal behavioral assessments: the Behavior Assessment System for Children, Second Edition, and the Achenbach System of Empirically Based Assessment. Both of these measures focus on emotions and behaviors. Finally, the chapter concludes with frequent errors that occur in behavioral assessment. These common errors are observer bias, observer drift, and observer fatigue.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 14

Assessment in Marriage and Family Counseling

Learning Objectives

- LO1 Describe the difficulties associated with developing measures for couples and families
- LO2 Understand how to conduct a family interview
- LO3 Explain how to conduct a genogram
- LO4 Describe the difficulties associated with using an instrument designed for individuals with couples and families
- LO5 Understand the constructs typically measured in couples' assessments
- LO6 Understand the constructs frequently assessed in family assessments

Rebecca called and requested family counseling for herself and her two sons, Zack and Josh. Rebecca started the counseling session by explaining that her husband, Ben, had recently moved into an apartment; however, the reason she sought family counseling was to address Zack's hostility toward her and his younger brother, Josh. Rebecca explained that Zack is 15 years old and, in the past six months, has become sullen and withdrawn. Furthermore, according to Rebecca, when Zack is asked to complete any household duties or attend to his homework, he becomes openly hostile to her and refuses to do anything she asks him to do. Rebecca also indicated that Zack seems to provoke verbal arguments with Josh, which had escalated into two physical altercations between the brothers. Josh is two years younger than Zack, and Rebecca is concerned that Zack might hurt his younger brother if they continue to fight. As their counselor, you asked Zack, "What do you think about what your mother was saying?" He responded by saying, "I don't know." You then asked Josh to describe the fights between him and Zack, and he replied, "They were no big deal." Both Zack and Josh appeared to have little interest in talking and stared at their shoes during the entire initial session. You begin to sense some kind of undercurrent within this family, but it is difficult to assess because neither of the boys are willing to respond with more than a word or two to any comments or questions. Hence, you would like another way of gathering information about the dynamics of this family and issues that are related to their interactions.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Thus far, this section of the book has focused on methods for assessing individuals; however, many counselors work with couples and families. Family relationships can be the sources of the greatest joys in life, but they can also be the cause of the greatest distress and sorrow when things go badly (Thomlison, 2010). Many marriage and family therapists or counselors use a systemic approach when working with couples or families. A systemic approach is holistic with a focus on the interactions that occur and the belief that all members of the family interact and influence each other. According to this philosophical view, assessment would not be of individuals but would focus on the dynamics or patterns that occur within the couple or family. Thus, testing each member within the family individually is not consistent with this systemic view.

Cierpka (2005a) extends this notion of clinical family assessment examining the family system and relational connections including the therapist's interactions within the family. Cierpka also stressed that family assessment is not done once to identify "the problem" but rather the assessment is of the current relational status and the assessment process is extended with attention to changes that may be occurring in the family.

Assessment of couples and families has emanated from many different fields, such as anthropology, counseling, human development, nursing, psychology, social work, and sociology. Within family therapy there are several different approaches: (a) behavioral, (b) systems and interactional, (c) interactional, and (d) psychoanalytic and objects relations. Discussion of family therapy theoretical perspectives is beyond the scope of this book, but students interested in couples and family assessment need training and supervision related to working with couples and families. Central, however, to all of these approaches is the belief that when change is needed, instead of isolating the individual from the emotional origins of conflict, the understanding of the network of family relationships and the family environment are central in conceptualizing the issues (Thomlison, 2010). It should also be noted that a family's culture, ethnicity, and religion have implications in the assessment of family systems.

Sometimes, counselors use instruments designed for individuals, such as personality instruments, when conducting marriage and family assessment. For example, some practitioners have used the Myers-Briggs Type Indicator (MBTI) to examine the differences in different family members' preferences. There are some difficulties, however, with using these tools because the assessment of individuals is essentially linear in concept—certain traits result in expected behaviors. Many marriage and family clinicians are grounded in *systemic theories.*

There are numerous measures for assessing couples and families, as evidenced by Touliatos, Perlmutter, and Straus (2001) compiling information on approximately 1,300 instruments. It does not appear that these instruments are being used in practice, however, because a survey of marriage and family therapists revealed that only 39% reported using standardized assessment instruments on a regular basis (Boughner, Hayes, Bubenzer, & West, 1994). This study also indicated that the marriage and family field does not use a specific battery of instruments because no single assessment instrument was used by more than 8% of the sample. An inspection of the instruments used by marriage and family therapists reflects that the majority of these instruments were designed for individual counseling as compared with couples and family counseling. One reason for the dearth of instruments designed for assessing couples and families is that the field of family therapy or counseling is a comparatively recent development. Before the 1950s, the dominant paradigm was individual therapeutic approaches; only in the past 60 years have theories or approaches for working with families been developed. As therapeutic approaches designed for working with couples and families evolved, so did the beginnings of formal assessment instruments in this area. Hence, formal instruments designed specifically for couples and families do not have the long history that, for example,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

personality measures have. Another reason for the dearth of instruments designed for assessing couples and families concerns the difficulties associated with developing sound instruments in this area.

Difficulties with Assessment of Couples and Families

L01 Numerous difficulties are associated with developing methods for assessing couples and families. One problem is in determining what should be assessed within families. Assuming a clinician adheres to a systemic approach, what then is meant by a family's or couple's dynamics? Even with specific terms, such as *interaction patterns*, there continue to be multiple views on what constitutes interaction patterns. Some argue that it is impossible to assess family interaction patterns because every family is so unique that these elusive dynamics cannot be quantified. Assuming a clinician does identify some dynamic or characteristic of a family, how do we then determine what constitutes a well-functioning family as opposed to a dysfunctional one? Let us take, for example, family cohesion and determining the amount of cohesion that is considered appropriate. Not all family theorists maintain that low cohesion is detrimental and high cohesion is positive in families. The Olson Circumplex Model (Olson, DeFrain, & Skogrand, 2010) suggests that cohesion should be viewed from a curvilinear perspective in which both too little and too much cohesion is problematic. In addition, what may be considered low cohesion for one family may function very well for another family. The difficulty in defining what constitutes functional as compared with dysfunctional becomes even more problematic when cultural differences are considered. Many differences exist between cultures regarding what are considered acceptable and unacceptable family behaviors. As an example, polygamy is accepted in some cultures and discouraged in others.

Family assessment is a broad term, because counselors might want very diverse information depending on the family's problems or issues. For example, a screening tool to determine if family counseling would be a suitable modality would be very different from an instrument to identify relational problems in a couple. Therefore, there is not just one area of focus in family assessment because the focus could be on couples' satisfaction, the identification of communication patterns, perceptions of the family environment, self-reports of behaviors (e.g., battering), an analysis of relationships, the identification of problems, or many other familial aspects. A family assessment device or technique could be extremely helpful in identifying the issues in one family but not appropriate for another family. Another problem with family assessment is that many of the variables assessed are fluid and fluctuate as compared to the more stable traits or characteristics of individuals. As an example of how family dynamics may vacillate, consider the variation in the level of satisfaction a couple might report depending on whether they spent the night before dining at a wonderful French restaurant or completing their income tax forms. In addition to situational fluctuations, there are also developmental changes. As Carter, McGoldrick, and Garcia-Preto (2010) have documented, families go through developmental stages within a life cycle. Certainly, a couple has different developmental issues when they have been married for four months as compared with their issues when they have been married for 21 years and have two adolescent children. Furthermore, individuals who are technically in the same group may be at different developmental levels. For example, some newlyweds may be 17 years old, whereas others may be in their late 30s. Therefore, the assessment process needs to consider the developmental stage or level of both the family and the individual members within the family.

Another issue in couple or family assessment concerns getting multiple views rather than a single client's perspective. Members within a family often have divergent views on

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the problems and issues, which can lead to a problem in the scoring of family assessments. Olson, Portner, and Lavee (1985) found low correlations between husbands' and wives' perceptions of family adaptability and cohesion. Thus, if the husband's perceptions are highly positive and the wife's very negative, should the scores be added together and then averaged? Interpreting the results from the different family members' perspectives is more difficult than interpreting a single client's results. Another problem with gathering multiple perspectives concerns the confidentiality of the results of a family assessment. Sometimes problems arise when family members have access to each other's results. As an example, a child might be reluctant to disclose some information if the parents are going to see the results. Thus, the counselor needs to consider how the assessment information is going to be used and who will have access to the results.

The final problem with couple and family assessment concerns the difficulties associated with developing adequate norming samples. It is a time-consuming and arduous task to build adequate norms for instruments designed for individuals; however, that task is small compared with instruments for couples and families in which the norms need to be made up of couples and families. An instrument developer must assess entire families rather than single individuals. Gathering assessment information from each member of a family is complex and problematic. Also, what is the norm for the couple or family, as only 25% of the population is in families made up of two opposite-gender married parents with two children, while the other 75% is made up of single parents, same-sex couples, children living with grandparents, grandparents living with adult children, or other configurations (L'Abate, 2012). The instrument developer should also consider the ethnic makeup of the norming group. As Morris (1990) pointed out, many of the participants in family assessment norming groups are primarily Anglo Americans. This overrepresentation of Anglo Americans is problematic because not all ethnic groups share the same perspectives of family. Cultural values have a notable influence on what is considered functional and dysfunctional in families. Therefore, it may not be appropriate to evaluate a family who recently immigrated to the United States from Afghanistan with an instrument that has a norming group from one Midwestern state that is 90% Anglo Americans. Although problems with norming groups can be overcome, developing an instrument with a large and representative sample is difficult.

Even though numerous difficulties are associated with the development of marriage and family assessment tools, a number of assessment devices have evolved that are departures from traditional paper-and-pencil instruments. Clinicians often interview families and couples, and this will be discussed first. The second type of family assessment is mapping techniques, and the most widely known of these assessment tools is the genogram. A third area of assessment in marriage and family counseling involves instruments designed for use in individual counseling. Many counselors use instruments designed for individuals and administer these instruments to every member of the family. A majority of these individually oriented instruments measure personality, although there are examples in the literature of counselors using the Block Design subscale of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) to examine differences in couples' task approach skills. The fourth area of assessment encompasses instruments designed specifically for assessment of couples and families. These instruments strive to assess the dynamics that occur among family members. The next four sections of this chapter address the multiple methods of assessing couples and families within the context of family interviews, mapping techniques, instruments designed for individual counseling, and instruments specifically developed for couples and family assessment. The separation of these approaches is quite the antithesis of how assessment is viewed in family therapy because assessment is not considered a distinct phase or process because the therapist blends various methods in attempting to assess the family system.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

LO2 Family Interviews

Thomlison (2010) suggested that one of the major modes of assessing families is through interviewing. In interviewing a family, the counselor needs to establish rapport and encourage family engagement. Family engagement is closely associated with joining the family or establishing a therapeutic alliance. Engagement provides the foundation for future interventions. Thomlison contended that there are many advantages to conducting the first interview in the family's home because the family may be more comfortable in their typical surroundings. Conducting the interview in the family's physical surrounding may also provide clinically relevant information to the clinicians, but clinician safety concerns should be addressed before attempting to interview a family in their home. Thomlison also advised clinicians to never interview a child or a victim of violence if the alleged perpetrator is in the room or even in the building.

Regarding interviewing families, it sometimes is important to note that several studies have found important difference among family members' reports of family functioning (Cook & Kenny, 2004). Therefore, the counselor should attempt to involve all members of the family in the interviewing process and observe all family members. Observation is often used in all types of assessment, but it is particularly important in the assessment of couples and families. To illustrate this point, you might consider how individuals in your family might complete an instrument concerning family dynamics as compared with what a skilled observer may see when examining your family's interactions. With a paper-and-pencil instrument, it is nearly impossible to measure the changes in tone, facial expressions, and nonverbal messages that often accompany verbal communications in families. On the other hand, observational approaches such as interviews also have problems, because counselors can misinterpret nonverbal expressions and inaccurately assess the family dynamics. To illustrate this point, let's examine the research related to differences in the amount of eye contact between distressed and nondistressed couples. Some clinicians might hypothesize that nondistressed couples are more likely to look into each other's eyes, but Haynes, Follingstad, and Sullivan (1979) found that distressed couples had a higher rate of eye contact than did nondistressed couples. Problems with observation are magnified as more clients are added to the process. A counselor cannot focus just on one individual in a couple or family. With this additional level of complexity, the reliability of the observations decreases. Furthermore, a counselor's biases and experiences both influence the inferences drawn from the family's behaviors. These inferences may be sound or distorted, depending on the counselor's ability to be objective.

Bray (2009) contended that there are six types of family characteristics that are frequently assessed in clinical evaluations of families: family structure and composition, family process, relationship patterns, family affect, family organization, and family diversity. *Family structure and composition* include family membership (e.g., couple only, couple with children, single-parent family) and the structure of the family (e.g., married couple, divorced family, stepfamily). *Family processes* go beyond content and include interactions and transactions among family members. *Relationship patterns* concern a sequences of couple or family interaction that develop over time, whereas *family affect* refers to the emotional tone and nature of the emotional expressions within the family. *Family organization* factors concern the roles and rules (both spoken and unspoken) within the family. It also concerns expectations for behavior that contribute to the family functioning. Finally, *family diversity* involves race/ethnicity, socioeconomic status, religion, and sexual orientation. Family diversity will influence all of the other family characteristics and the manifestations of them.

According to Duffy and Chenail (2012), some of the historical leaders in family therapy (e.g., Salvador Minuchin and Jay Haley) suggested that therapists observe the structure,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 14.1
Therapists' observations related to possible family structures, hierarchies, and interactional patterns

- Who sits close together and who sits away from whom?
- Who initiates conversations and who remains quiet during discussions of family issues?
- Who can interrupt or disagree and who appears unable to voice a differing opinion?
- Who identifies problems or goals for therapy?
- How are emotions expressed and responded to by family members?
- How are conflicts expressed and responded to by family members?
- Whose opinions seem to carry more weight and whose opinions are dismissed?
- What topics seem to be off limits for family members to discuss?
- Who attempts to solve problems and who is less involved?
- Who seems to team up with whom regarding taking positions?
- Who appears to be in charge of which activities and functions?
- Who is in charge of discipline and who is in charge of fun?
- Who has veto power over important decisions?
- Who can make independent decisions and who must be consulted?
- Who gets nurtured by whom and who gets less nurturing?

Adapted from Duffy and Chenail (2004).

hierarchy, and interactions of the family. Table 14.1 reflects some concepts for counselors to consider when they are observing families to gain insight into dynamics within the family. Some researchers in the area of marriage and family assessment have attempted to address the limitations in observation by developing coding systems (see Hampton & Beavers, 2012).

In conclusion, interviewing couples and families can often provide information about family characteristics and dynamics. Interviewing families is complex as the clinicians must not only ask clinically relevant questions but also observe all family members while conducting the interview. The interview, however, can often provide insights into family dynamics that may be corroborated by other assessment strategies.

Mapping Techniques

Mapping techniques are where the counselor has the family collaborate with the counselor in graphically representing some pertinent representation. The most widely used of the mapping techniques is the family genogram. Counselors use mapping techniques to gain information and also to observe the family dynamics while the mapping technique is being completed.

L03 Genograms

Genograms offer an assessment technique often incorporated into counseling with couples and families (Duffy & Chenail, 2012) and are probably the most well-known assessment procedure in family counseling. Counselors work with families to help them draw a family tree that records information about family members and their relationships over at least three generations (McGoldrick, Gerson, & Petry, 2008). Genograms are most closely associated with Bowen's family system theory (Bowen, 1978), although counselors with other theoretical orientations use them as well. Bowen contended that there are multigenerational transmissions of family patterns. Genograms can help identify which family patterns are being repeated by the family who is completing the genogram. With genograms, the clinician examines the previous generations and looks for patterns of functioning,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

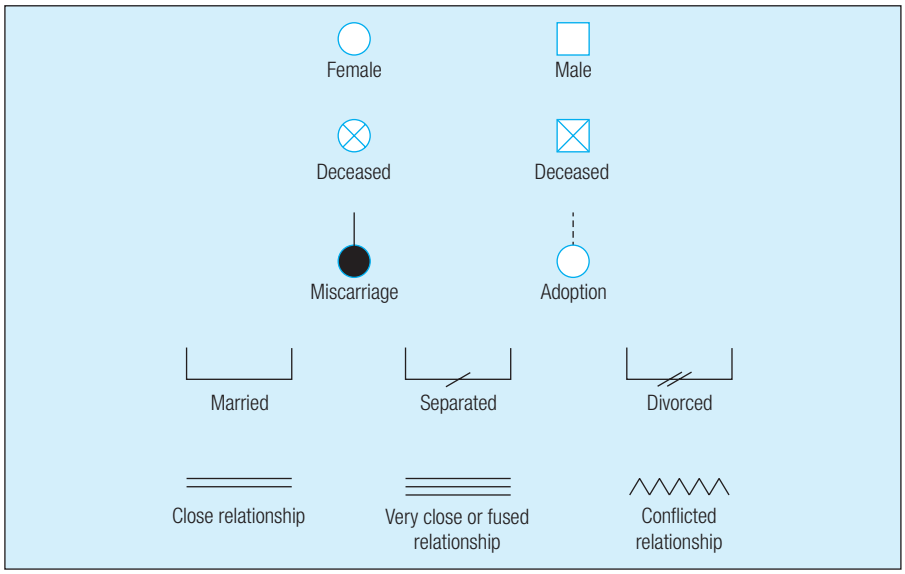
relationships, and structure. A basic assumption with the use of genograms is that problems and symptoms reflect a system's adaptation and reaction to its context.

According to McGoldrick et al. (2008), gathering family information and constructing a genogram should be integrated into the more general process of joining, assessing, and helping a family. The first step in creating a genogram involves mapping the family structure, which serves as the skeletal structure or backbone of the genogram. Typically, the counselor asks questions about family structure and maps the genogram for the past three generations. The mapping is a graphic depiction of how family members are biologically, legally, and emotionally related to one another over three generations (McGoldrick et al., 2008). This mapping involves circles and squares that represent people and lines that symbolize their relationships.

Circles represent female family members; squares, males. Figure 14.1 provides examples of many of the symbols used in a genogram, and Figure 14.2 is an example of an actual genogram. Horizontal lines connecting two symbols indicate married individuals. If the couple is a love affair or living together, then the line is a series of dashes as compared to solid to indicate the relationship is not legal. If they are in a committed relationship, then the relationship is signified by both a dotted line and a solid line. Separations are indicated by one slashed line, and divorce by two slashed lines. All pregnancies, stillbirths, miscarriages, and abortions are mapped. If an individual has died, then an X is drawn through the corresponding circle or square.

In Figure 14.2, the genogram represents a couple, Jody and Brad, who sought counseling because they are considering getting married but are concerned about the recent strife within the relationship. As the genogram depicts, Jody and Brad has been living together for two years in a committed relationship. On Jody's side of the genogram, her parents are Richard and Elizabeth who were divorced in 1994, when Jody was 11 years old. Jody has one brother, Evan, and a half-brother, Noah. Jody's mother is currently married to Todd, who is the father of Noah. Jody's father's parents are Eileen and Frank and Jody's father was their only child. Jody's grandparents on her mother's side of the family are Marianne and Robert, who had three children: Jon, Elizabeth (Jody's mother), and Kate. On Brad's side of the genogram, he is the only child of Bill and Irene. Brad's father's parents are Henry and Mary, who had three children: Bill (Brad's father), Cathy, and Mike. Brad's mother, Irene, is the only child of Ed and Peggy.

FIGURE 14.1
Symbols used in a genogram



Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

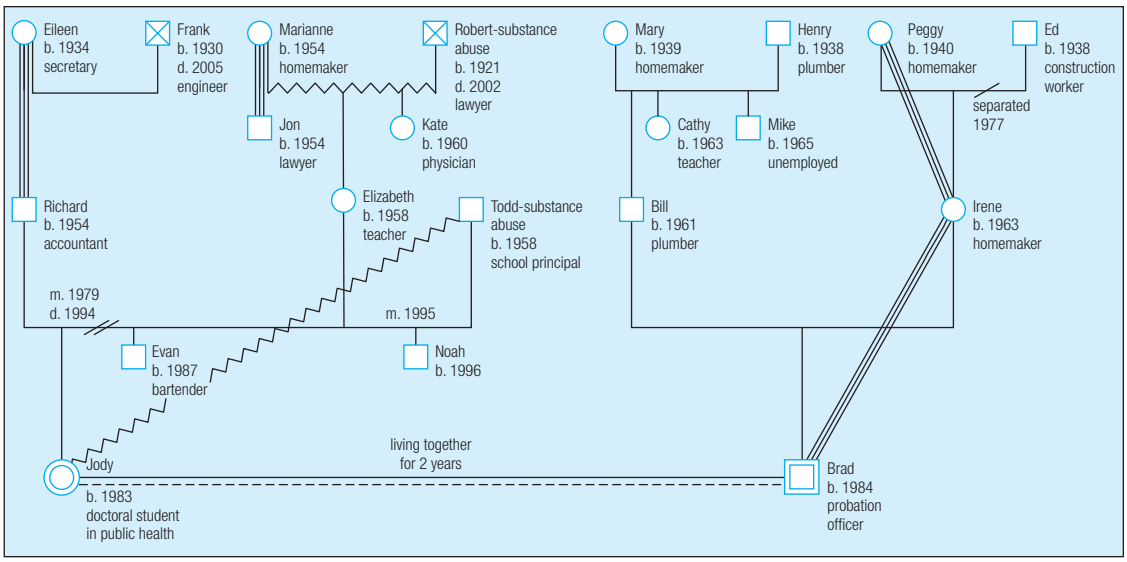


FIGURE 14.2
Example of a genogram

Once the skeleton of the genogram is assembled, the next step generally is to record family information, which can include demographic information, functioning information, and critical family events. Demographic information includes marking the year of each individual's birth (e.g., b. 1983) and the year of death (e.g., d. 1992). Dates of marriages and divorces are also recorded on the genogram. As Figure 14.2 shows, Jody's parents were married in 1979 and divorced in 1994. The genogram also indicates that Brad's grandparents, although they never divorced, separated in 1977. Often during this stage in the genogram construction, which McGoldrick et al. (2008) called "getting the facts," the clinician will also note information related to occupations, education, religious practices, or current situations. For example, as can be seen in Figure 14.2, Brad's mother and both grandmothers were all homemakers. In mapping the genogram, this fact led to a discussion that Jody thinks Brad wants her to do all of the housework as Brad's mother and grandmothers seemed to do. Another occupational theme is that on Jody's side of the family, the men tend to be employed in professional occupations (e.g., lawyers and an accountant); whereas, the men from Brad's side of the family tend to be involved in construction and plumbing. This difference in family education may also be reflected in Jody and Brad's occupational choices as Jody is working on a doctoral degree in public health and Brad has a bachelor's degree in criminal justices.

After mapping the family's structure, the clinician begins to gather information related to patterns of functioning. Critical family events, such as losses, successes, or transitions, are also recorded. If an event is deemed particularly noteworthy, a counselor may map it in the genogram and note it again in the margins. The theoretical foundation of genograms is that family patterns are repeated; thus, the clinician gathers insights into current problems by examining previous family functioning. Included in Figure 14.1 are lines for mapping relational dimensions, such as close, very close or fused, and conflicted relationships. An examination of Figure 14.2 shows that there are a number of close or very close relationships. For example, Brad has a close, possibly fused, relationship with his mother, Irene. His mother also had a very close or fused relationship with her mother. Jody is also familiar with very close relationship as her father had one with his mother, and her grandmother, Marianne, had one with her Uncle Jon. This familiarity with close or

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

possibly enmeshed relationship may contribute to Jody's discomfort with the degree of closeness between Brad and his mother. Another possible transgenerational trend may be related to substance abuse and, in particular, the tendency for women in Jody's family to marry men with substance abuse issues. If you examine Figure 14.2, you may notice that both Jody's maternal grandmother and her mother married men with substance abuse issues. Therefore, the counselor may want to explore Brad's substance use habits to see if this trend has continued. You may also want to examine Figure 14.2 or map your own genogram and consider family themes.

Interpreting the genogram involves constructing hypotheses that are later examined with the family. The counselor examines the relational structure and variables, such as whether the composition of the household is a nuclear family, the involvement of extended family, the order of the siblings, and any unusual family configurations. The family life cycle and stage should also be considered in the interpretation. As indicated earlier, genograms are based on the premise that family patterns can be transmitted from one generation to the next and by recognizing these patterns, the counselor can assist the family in avoiding ineffectual patterns. The genogram can also be used to analyze systemic connections among critical events that, at first, may not appear interrelated. Family therapists often focus on triangular relationships or triangles. The formation of triangles in families typically involves two people "triangulating" or bringing a third member into their relationship, which usually serves to lessen the difficulties in the initial dyad. For example, two members of a family may join together to help a third member, or two members could join and gang up against a third. According to Bowen (1978), it is the collusion of the two in relation to the third that is the defining aspect of a triangle. In a general sense, the genogram is designed to provide information on the family structure, roles, level of functioning, and resources. However, the counselor works together with the family to create the genogram, and the clinician observes the family as the information is recorded. Both the observation and the recorded information are used in interpreting the genogram. The final form a genogram takes varies depending on how much detail is gathered and the patterns being identified. As an example, therapists would take more time examining a history of sexual abuse (Timm & Blow, 2005). A counselor should also consider issues of culture and racial/ethnic differences that may influence familial patterns.

Some psychometric limitations need to be considered before using a genogram. Timm and Blow (2005) contended that there have been very few studies on the psychometrics of the genogram and that there had been only study on its reliability (Coupland, Serovic, & Glenn, 1995). Rohrbraugh, Rogers, and McGoldrick (1992) found poor interjudge reliability among practitioners in a medical setting. Coupland et al. found that doctoral students varied greatly in their accuracy in recording family information on genograms. This finding is even a little more disturbing because these advanced students had been trained in how to construct genograms. It also should be stressed that there is very little validation evidence for genograms. Therefore, clinicians should be extremely cautious in the conclusions they reach based on genogram information. Similar to other observational techniques, genograms are particularly susceptible to counselor bias. Counselors need to view their interpretations of genograms as only hypotheses that need to be confirmed or invalidated by gathering additional information about the family.

Family Sculpting

Another mapping technique that has historically been associated with family therapy is family sculpting. Family sculpting emerged in the late 1960s as a method to assess and intervene with families (Bischof & Helmeke, 2005). The technique is associated with the work of Virginia Satir. Family sculpture involves positioning either family

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

members or objects that represent the family in physical space that symbolizes the relations and roles within the family system. The steps in conducting a family sculpture, described in Table 14.2, are adapted from Bischof and Helmeke (2005). Sometimes after completing one family sculpture, the counselor will begin another with a different member of the family. This can facilitate an exploration of different perspectives within the family.

There is considerable subjectivity and inconsistency in family sculpture, but Cromwell, Fournier, and Kvebaek (1980) developed the *Kvebaek Family Sculpture Technique*, which is more structured than most sculpting techniques. This technique uses a board that is similar to a chessboard except there are 100 squares. The family members are instructed first individually and then as a family to place figurines on the board depicting the relationships among the family. The family first represents the relationships as they currently exist and then sculpts the family again from an ideal perspective. The counselor assesses the sculptures by measuring the distances between the various placements of the figurines. Marriage and family therapists use sculpting techniques with a variety of objects, including family members arranging themselves in the room to depict their relationships. It is important to remember that sculpting techniques tend to be unreliable. Relationships within a family are fluid, and these techniques reflect only the clients' current perspectives. Therefore, counselors need to be cautious in interpreting this information and should not consider it a stable representation of the family.

TABLE 14.2
Steps in conducting a family sculpture

- **Introduction and Warm-Up of Family**
  The counselor introduces the sculpting exercise, describes the process, and provides the rationale for using the sculpting technique.
- **Selecting a Sculptor**
  The therapist either specifically selects an individual to be the sculptor or asks for a volunteer.
- **Defining the Focus**
  The counselor helps focus the sculpture and may select a specific time in the family's history, a specific family situation or problem, or specific themes within the family (e.g., power, relationships).
- **Placing Family Members in the Sculpture**
  In this step, the family sculptor places family members physically as a representation of the focus. The therapist establishes that there should be little talking, and family members may not challenge the instructions of the sculptor.
- **Detailing the Sculpture**
  Either after each person is placed or after the entire sculpture is completed, the counselor helps the sculptor express details about the placement. This is the period when the therapist may explore clinical hypotheses about the interactions in the family.
- **Sculptor Adding Self**
  Once the family members are all placed and the sculpture has been adequately explored, the sculptor adds himself or herself. Because this may influence the perspective of the sculptor, additional adjustments can be made. The counselor may also help sharpen the description of the sculpture.
- **Choice Point**
  At this point, the therapist needs to make some clinical decisions regarding the therapeutic theme of the sculpture. The therapist might give the sculpture a name, describe the patterns that have emerged, or have different family members explore embedded patterns.
- **Deroling and Debriefing with the Sculptor and the Participants**
  In this step, the counselor carefully has the sculptor and the family members leave their roles in the sculpture and return to the present. Often the therapist's theoretical approach will influence this process.

Adapted from Bischof and Helmeke (2005).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Other Mapping Techniques

L'Abate (1994) suggested that the evaluation of a family should involve a battery of techniques, which might include a genogram and other mapping techniques. One of the mapping techniques described by L'Abate is the *ecomap*, which is used to gather information about the larger context of the family and the home. Gathering this information involves a collaborative process between the counselor and all members of the family. The process starts by positioning the home in the center of a piece of paper. Each family member then draws how far and where other resources are that are necessary for his or her "survival." The family members typically map schools, grocery stores, shopping, church, and recreational activities. The counselor can ask questions as the family creates the ecomap. The process the family uses to complete the ecomap may provide some indications of the dynamics in the family (e.g., Does one member dominate? Does one member not participate?). For more information on how to conduct an ecomap see L'Abate, Cusinato, Maino, Colesso, and Scilletta (2010).

In conclusion, mapping and observation activities should not be used as the sole means of family assessment because of the problems with reliability and the limited validation information. When using a more subjective assessment tool, the counselor needs to balance the process with other, more objective techniques. As is often repeated in this book, multiple forms of assessment are preferred because this provides a more thorough picture of the client or, in the case of couple and family assessment, the clients.

LO4 Use of Instruments Designed for Individuals

A number of instruments initially constructed for use in individual counseling have been adapted for use in couple and family assessment. Boughner et al. (1994) documented the prevalence of using an assessment instrument for individuals in marriage and family therapy. They found that the most frequently used assessment instruments in marriage and family therapy are the *Myers-Briggs Type Indicator*, the *Minnesota Multiphasic Personality Inventory-2* (MMPI-2), and the *Taylor-Johnson Temperament Analysis*. Given the interest in therapeutic interventions with couples and families, two new scales have been added to the scoring of the MMPI-2 (Nurse & Sperry, 2012). These scales are the Family Problem Scale (FAM) and the Marital Distress Scale (MDS). Other individually oriented instruments that marriage and family therapists reported using were the 16 Personality Factor Questionnaire (16PF), the Millon Multiaxial Inventory, and the Beck Depression Inventory. In using these individually oriented instruments, the counselor, working with a family, often uses the information to identify how differences in personality or preferences may contribute to family issues. For counselors interested in using personality assessments with couples and families, Nurse (1999) provided some excellent examples and insights into the process.

Cierpka (2005b) proposed a three-level model of family assessment in which distinctions are made between the individual, the dyadic or triadic levels, and the system as a whole. In this case, an assessment of an individual could occur, but the focus of gathering this information is on how the individual organizes himself or herself within the family. I will use the MBTI as an example of how an individually oriented assessment can be used to gain information on how the individual organizes herself or himself within the family. Rather than focusing on each individual's results, with a family, the focus would be on the interactions that occur using the framework of the MBTI. As an illustration, let's say that a counselor is conducting couples counseling and the wife's MBTI code is ENFP and the husband's code is ISTJ. The counselor could explain to the couple that in times of stress, the wife, who has the extroversion (E) preference, will seek others. The husband, on the other hand, with the introversion (I) preference, will desire time to

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

himself. Hence, the interaction between them is difficult as they seek different methods for coping with stress. Furthermore, the couple's preferences on the second dimension, which concerns how they perceive information, may also contribute to relational difficulties. The husband prefers to use his senses (S), and he may become irritated that his wife misses what he considers important details. His wife, on the other hand, who prefers to use intuition (N) in perceiving information, may be annoyed because her husband does not perceive the "big picture" and is more interested in insignificant details. This couple also varies on the third dimension, which concerns how they make sense of the information they perceive. The wife prefers to use her feelings (F), as compared to the husband, who prefers thinking (T). In this case, the husband may think the wife should be more rational and logical in her decision making, whereas the wife may perceive the husband to be cold and uncaring. The fourth dimension of the MBTI concerns differences in approaches to the outer world and whether individuals prefer to have their lives planned and organized or whether they prefer to be more spontaneous and respond to the moment. Once again, the couple's profiles indicate that there are differences. The wife prefers to react spontaneously, and the husband prefers to have activities arranged and organized. Thus, the marital difficulties are not related to one person's problem or flawed personality, but the focus of counseling is on the fact that each individual has different preferences in perceiving and processing information. One reason for the popularity of the MBTI with couples and families is it does not label preferences as right and wrong or good and bad. Instead, the instrument is designed to understand and respect differences.

Another instrument originally designed for use in individual counseling that is now frequently used in counseling couples and families is the *Taylor-Johnson Temperament Analysis* (T-JTA). The T-JTA has new norms that were published in 2007 and a new manual that also was released in 2007. The Taylor-Johnson Temperament Analysis has been used for more than 40 years in individual, premarital, marriage, and family counseling. This tool was originally designed as a measure of an individual client's personality. The authors later developed a feature called "criss-cross" testing, which can be particularly useful in counseling couples. With the criss-cross technique, each individual first takes the instrument. The individuals then take it a second time, but this time they answer as they believe the other person would answer. Thus, the profile contains the individual's self-appraisal and the other person's perceptions of that individual. In marital and premarital counseling, the counselor then can use the criss-cross testing results to look at congruence in the couple's personalities as well as at congruence in how individuals view themselves and how their spouses view them.

Some clinicians, in the past, have objected to the profile sheets of the T-JTA because the sheets contained four shaded areas that corresponded to certain percentile scores. These shaded areas were labeled with the clinical designations of Excellent, Acceptable, Improvement Desirable, and Improvement Needed. Due to the complaints about the labels, it is now possible to purchase profile sheets with or without these clinical designations. Another previous problem with the T-JTA was an inadequate norming sample; however, the current version of the instrument was normed on 27,000 individuals.

Formalized Assessment Instruments Specific to Marriage and Family

Fredman and Sherman (1987) asserted that some of these well-known individually oriented instruments can provide therapists with insights into the family. These instruments do not, however, provide insight into level of satisfaction, qualities of the relationship, areas of conflict, or measures of parenting skill. During the past 50 years, researchers have studied marriages and families (Touliatos et al., 2001); however, this research on dynamics among couples and families is not reflected in instruments designed for individuals. Furthermore,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

many clinicians working with couples and families use a systemic approach, and although individually oriented instruments may describe the individuals within the system, these tools do not describe the family system. Hence, in the past 20 years, there has been significant interest in instruments designed specifically for assessing systemic family factors. This section focuses on assessment instruments designed specifically for use with couples and families. The first area addressed is measures designed expressly for premarital counseling, followed by couples' assessment, and, finally, instruments used to assess families.

L05 Assessment in Premarital Counseling

In the area of premarital counseling, Boughner et al. (1994) found that practitioners reported using the *PREmarital Personal and Relationship Evaluation* (PREPARE) most often. In addition, PREPARE is one of the three premarital assessment questionnaires endorsed by the American Association of Marriage and Family Therapists (Larson, 2002). PREPARE is one of five instruments of the PREPARE/ENRICH program: (1) PREPARE for premarital couples, (2) PREPARE-MC for premarital couples with children, (3) PREPARE-CC for cohabiting premarital couples with or without children, (4) ENRICH for married couples, and (5) MATE for couples over the age of 50.

The entire program is designed to have the couples work with a trained counselor to (1) explore relationship strengths and limitations, (2) learn assertiveness and active listening skills, (3) learn how to resolve conflict using the 10-step model, (4) help the couple discuss their families of origin, and (5) help the couple with financial planning and budgeting. Couples are assessed in 20 areas, with 13 scales relating to significant couples issues (e.g., communication, conflict resolution, personality issues, financial management, marital satisfaction, and leisure activities). Another four scales measure personality in the areas of assertiveness, self-confidence, avoidance, and partner dominance, and two scales describe family of origin and the type of marriage/couple relationships in terms of closeness and flexibility.

In a study of PREPARE (Version 2000), Knutson and Olson (2003) found that those couples who participated in the PREPARE program and its feedback group reported greater couple satisfaction after participating than did couples who only completed PREPARE but received no feedback and a control group that did not take PREPARE. Futris, Barton, Aholou, and Seponski (2011) found that engaged couples who participated in a PREPARE program reported positive gains in knowledge, felt more confident in their relationship, exhibited more conflict management strategies, and reported feeling more confident in their relationship. Also, these authors found that there was not a difference between those who attended a one-day workshop and those who attended a series of sessions. In order to use PREPARE, you must receive training, which can be done through a workshop or online.

Another assessment instrument for engaged couples endorsed by the American Association of Marriage and Family Therapists (Larson, 2002) is RELATE, which was developed by the Marriage Study Consortium at Brigham Young University. Although RELATE started as an assessment for engaged couples, it is now available for any couple in a committed relationship. RELATE is available over the Internet (relate-institute.org), and clients can go directly to the website to take the instrument. According to Busby, Holman, and Taniguchi (2001), RELATE's conceptual model was drawn from more than 50 years of research regarding premarital predictors of later marital quality and stability. RELATE is comprised of 276 items that measure the four broad areas of personality/values, family and friend support, communication skills, and upbringing/background. Counselors can also set up accounts and have their clients take the assessments online. In a randomized controlled trial study, Larson, Vatter, Galbraith, Holman, and Stahmann (2007) found

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

that feedback using the Internet-based RELATE inventory increased relationship satisfaction relative to a wait list control. Furthermore, a study by Busby et al. (2001) examined some of the basic psychometric qualities of the instrument.

Couples or Marital Assessment

According to Robert and Robins (2000), individuals rate having a satisfying marriage or relationship as one of the most important goals in their lives. In assessing couples, the focus tends to be on the relationship and couple distress (Snyder, Heyman, & Hayes, 2009). One of the first widely used assessment instruments for couples was the *Marital Adjustment Test* (Locke & Wallace, 1959), which was often referred to as the Locke-Wallace. This instrument has been used for validating many of the later marital scales. As an indicator of where the field was at this point, the Marital Adjustment Test contained only 15 items, and the norming group was 118 husbands and 118 wives who were *not* necessarily married to each other.

Dyadic Adjustment Scale. The *Dyadic Adjustment Scale* (DAS; Spanier, 1976) is another instrument often used in marital counseling. It is intended to measure the adjustment quality of married couples or similar dyads. This 32-item instrument produces a total adjustment score plus scores on four subscales: Dyadic Consensus, Dyadic Satisfaction, Dyadic Cohesion, and Affectional Expression. The overall reliability of this instrument is quite high (.96), but the norming group consists of only 218 white married individuals from Pennsylvania. Graham, Liu, and Jeziorski (2006) contended that the DAS is probably the most widely used assessment of relationship quality. They found, using reliability generalization analyses, that the DAS total score's average reliability coefficient is .91 and the subscale averages range from .71 to .87. Supporting the contention that the DAS can be used with a wide range of couples, they found that reliability estimates did not differ based on sexual orientation, gender, marital status, or ethnicity of the samples. There is also a short form of the DAS that can be used to monitor treatment. The importance of monitoring treatment is discussed in Chapter 16.

Marital Satisfaction Inventory-Revised. One of the most widely used instrument in couples counseling is the *Marital Satisfaction Inventory-Revised* (MSI-R; Snyder, 1997), which was designed to assist couples in communicating about a broad range of relationship issues. In the revised edition, to be appropriate for both traditional and nontraditional couples, some of the items were changed to use terms such as *partner* and *relationship* rather than *spouse* and *marriage*. The items are scored on 13 scales (see Figure 14.3). The first scale in Figure 14.3 is Inconsistency (INC), which indicates the accuracy of the profile by showing the degree to which the clients answered certain questions in a consistent manner. Conventionalization (CNV) is the validity scale and reflects an individual's tendency to distort the appraisal of the marriage in a socially desirable manner.

The Global Distress Scale (GDS) is an overall measure of dissatisfaction. High scores on the MSI-R indicate dissatisfaction. The other 10 scales measure specific sources of conflict: Affective Communication (AFC), Problem-Solving Communication (PSC), Aggression (AGG), Time Together (TTO), Disagreement About Finances (FIN), Sexual Dissatisfaction (SEX), Role Orientation (ROR), Family History of Distress (FAM), Dissatisfaction with Children (DSC), and Conflict Over Child Rearing (CCR). The items for the last two scales are answered only if the couple has children. The instrument was restandardized on 2,040 people (1,020 intact couples). The ages of the couples were from the late teens through the early 90s, and attempts were made to stratify the sample to approximate the U.S. population with regard to geographic region, education, and ethnicity.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The reliability coefficients for the scales range between .80 and .97. The manual contains validity information, which is meager when compared with personality instruments but substantial when compared with other marital assessments. Bernt (2001) contended that the MSI-R is particularly well suited to clinical settings and that the manual provides an extensive discussion on how to interpret the scales and use the information in counseling with couples.

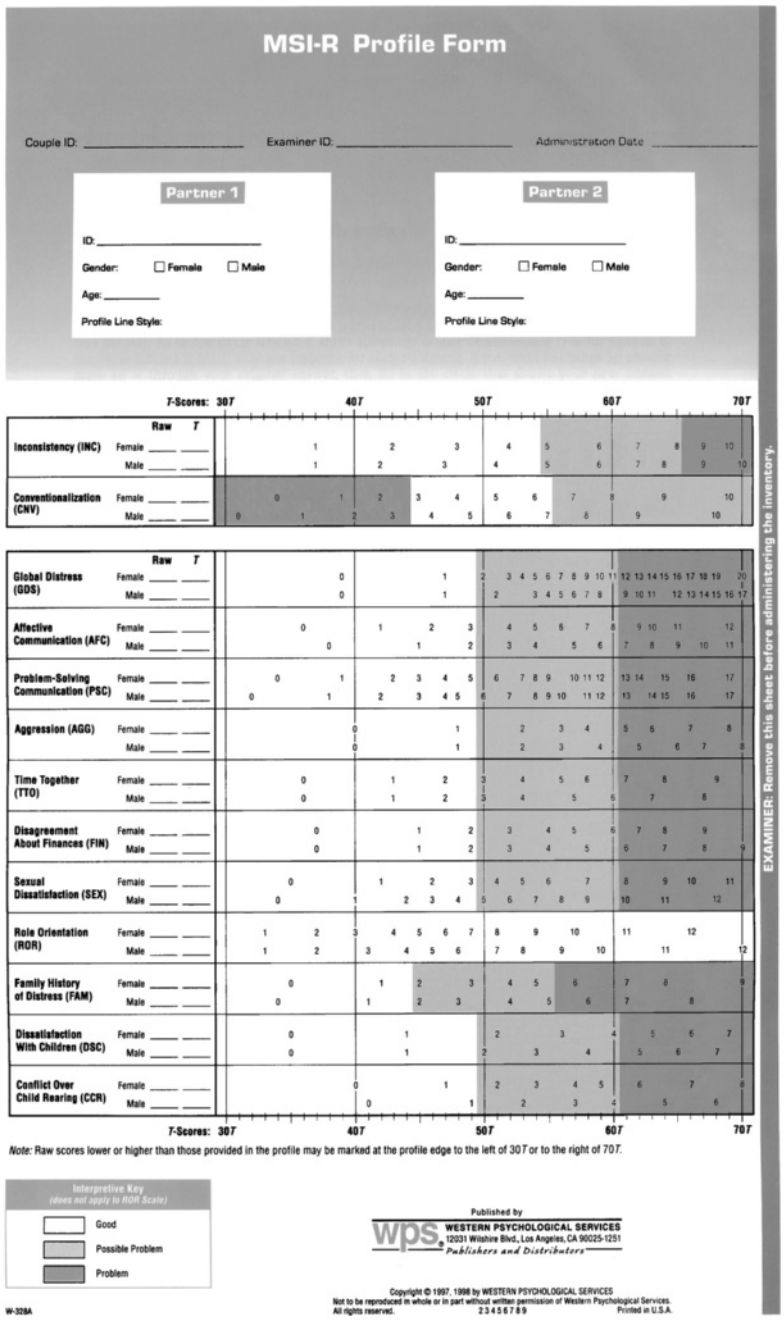

MSI-R Profile Form
Couple ID:
Examiner ID:
Administration Date
Partner 1
ID:
Gender: Female Male
Age:
Profile Line Style:
Partner 2
ID:
Gender: Female Male
Age:
Profile Line Style:
T-Scores: 30T 40T 50T 60T 70T
Raw T
Inconsistency (INC)
Conventionalization (CNV)
Global Distress (GDS)
Affective Communication (AFC)
Problem-Solving Communication (PSC)
Aggression (AGG)
Time Together (TTO)
Disagreement About Finances (FIN)
Sexual Dissatisfaction (SEX)
Role Orientation (ROR)
Family History of Distress (FAM)
Dissatisfaction With Children (DSC)
Conflict Over Child Rearing (CCR)
Female
Male
Note: Raw scores lower or higher than those provided in the profile may be marked at the profile edge to the left of 30T or to the right of 70T.
Interpretive Key
(does not apply to ROR Scale)
Good
Possible Problem
Problem
Published by
WESTERN PSYCHOLOGICAL SERVICES
12031 Wilshire Blvd., Los Angeles, CA 90025-1251
Publishers and Distributors
Copyright © 1997, 1998 by WESTERN PSYCHOLOGICAL SERVICES
Not to be reproduced in whole or in part without written permission of Western Psychological Services.
All rights reserved. 2 3 4 5 6 7 8 9 Printed in U.S.A.
W-328A
EXAMINER: Remove this sheet before administering the inventory.

FIGURE 14.3
Marital Satisfaction Inventory-Revised Profile Form. Material from the MSR-I copyright © 1997, 1998 by Western Psychological Services. Reprinted for the sole purpose of professional reference by Cengage Learning, by permission of WPS, 625 Alaska Avenue, Torrance, California, 90503, USA, rights@wpspublish.com. Not to be reprinted in whole or in part for any additional purpose without the expressed, written permission of the publisher. All rights reserved.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Marital Disaffection Scale. Kayser (1993) published the Marital Disaffection Scale, which measures the gradual loss of positive emotion (e.g., love, caring, closeness). Disaffection does not necessarily lead to the dissolution of the relationship as some couple may stay together even though there is a loss of positive emotion. The Marital Disaffection Scale is a 21-item instrument that can be taken in three to five minutes. Kayser (1993) reported a coefficient alpha of .93 and also provides some validation evidence. Although there are subscales (e.g., Attachment, Emotional Estrangement, and Emotional Support), Kayser recommends only using the total score.

LO6 Family Assessment Instruments

Family Adaptability and Cohesion Evaluation Scale IV. The *Family Adaptability and Cohesion Evaluation Scale IV* (FACES IV; Olson, Gorall, & Tiesel, 2004) is the latest version of a widely used measure of family dynamics (i.e., FACES-III). FACES IV was developed to measure the Circumplex Model of family functioning. The theorists asserted that the cohesion and flexibility dimensions are curvilinear with too little or too much being problematic. Family cohesion concerns the emotional bonding that couples and family members have toward one another. The definition of family flexibility is somewhat different from that given in earlier versions of FACES. Olson and Gorall (2006) defined *family flexibility* as "the quality and expression of leadership and organization, role relationships, and relationships rules" (p. 6). Two scales on FACES IV measure the degree to which cohesion and flexibility are balanced (i.e., Balanced Cohesion and Balanced Flexibility), with higher scores reflecting a good balance and healthy functioning. There are also four "unbalanced" scales that measure more problematic family function. Two of these unbalanced scales concern problems with cohesion: Disengaged Scale and Enmeshed Scale. There are also two unbalanced scales related to problems with flexibility: Chaotic Scale and Rigid Scale. The scores on these scales result in a family being categorized into one of six family types. The six family types are on a continuum from the most healthy and happy to the least healthy and most problematic and they are: Balanced, Rigidly Cohesive, Midrange, Flexibly Unbalanced, Chaotically Disengaged, and Unbalanced. Individuals interested in a detailed discussion of these six types of families may go to http://www.facesiv.com.

Family Assessment Measure. An assessment that appears to be gaining recognition is the *Family Assessment Measure* (FAM-III) by Skinner, Steinhauer, and Santa-Barbara (1995). In fact, the National Institute on Drug Abuse used the FAM-III as part of its comprehensive assessment battery of problem adolescents. The FAM-III is designed to assess family functioning that integrates individual members and collective characteristics of the family. Spillane (2001) contended that it is particularly good at probing characteristics and processes of families. The FAM-III consists of three interrelated instruments. The first is the General Scale (50 items), which focuses on the family's general functioning. The second instrument, the Dyadic Relationship Scale (42 items), measures the quality of the relationship between two specific family members (e.g., mother and father). The third scale is the Self-Rating Scale (42 items), in which individuals evaluate their own level of functioning in the family unit.

Each of the three scales provides information on the following seven dimensions: Task Accomplishment, Role Performance, Communication, Affective Expression, Involvement, Control, and Values and Norms. The FAM-III is available in both paper-and-pencil and computer formats. Manges (2001) found the FAM-III to be a valuable, but time-consuming, tool. The reported Cronbach's alpha for each scale is comparatively high for both adults and children, with the range for adults being between .89 and .95 and that for children being

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

from .86 to .94. The validity information is relatively limited but is growing rapidly. There is also a computerized version, in which the software scores and writes an interpretative report.

Family Environment Scale. The *Family Environment Scale* (FES; Moos & Moos, 1994) evolved out of Rudolf Moos's work on assessing social climates. Assessment of social climates is closely related to environmental assessment, which examines how environmental factors influence an individual's behavior. Moos (1973) contended that three dimensions provide insight into a social climate or environment: (1) a relationship dimension, (2) a personal development or personal growth dimension, and (3) a system maintenance and system change dimension. According to Moos, these dimensions provide insight into family environments as well as into other social climates or environments.

Table 14.3 describes the subscales related to each of the three dimensions in the FES. There are three forms of the FES: the Real Form (Form R) measures individuals' perception of their current family, the Ideal Form (Form I) measures preferences for an ideal family environment, and the Expectation Form (Form E) measures expectations about a family environment (e.g., prospective foster parents' expectations). Each form of the FES contains 90 items. These three forms can be used either individually or in combination in a variety of counseling interventions. Often, each family member's results are compared to examine differences in perceptions. For example, the counselor might want to explore discrepancies among the family members' perceptions of the environment. I once worked with a family in which the mother and the children rated the conflict high, but the father rated the conflict below average. Analysis of this discrepancy led to a useful discussion of the father's attempt to dismiss any opinion that varied from his stance on an issue. Another way to use the FES to explore family perceptions is to examine the differences between the results from the Real Form and those of the Ideal Form. Sometimes, a family will have amazing similarities among themselves on results from the Ideal Form, which can be an effective intervention for unifying the direction of the counseling.

The normative sample of the FES is 1,432 normal families and 788 distressed families. The manual does not provide very detailed information on the manner in which many of

TABLE 14.3
Dimensions and subscales of the family environment scale

Relationship Dimension

Cohesion—the amount of commitment, help, and support family members provide for each other

Expressive—the degree to which members of the family are encouraged to express feelings and act in an open manner

Conflict—the amount of openly expressed conflict and anger in the family

Personal Growth Dimension

Independence—the extent to which members are assertive, self-sufficient, and able to make their own decisions

Achievement Orientation—the degree to which activities are framed in an achievement-oriented or competitive framework

Intellectual-Cultural Orientation—the extent to which the family is interested and participates in political, intellectual, and cultural activities

Active-Recreational Orientation—the level of the family's participation in social and recreational activities

Moral-Religious Emphasis—the amount of emphasis on values and religious and ethical issues

System Maintenance Dimension

Organization—the extent to which organization and structure are clear in responsibilities and planning family activities

Control—the degree to which set rules and procedures run family life

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

the normal families were selected. The internal consistency coefficients range from .61 to .78 on the ten subscales of the FES. There is also considerable validity information on this instrument, much of which concerns differences between normal and distressed families.

Other Instruments. As indicated by Touliatos et al. (2001), hundreds of other family instruments are available to measure various familial aspects, such as family beliefs, the extent to which families celebrate, coping behaviors, and family empowerment. In selecting assessments to assess either couples or families, counselors need to examine the family constructs or dynamics that they are attempting to assess (e.g., hostility, communication patterns, domestic violence). Many of the instruments in the marriage and family area have limited psychometric information; thus, clinicians should be cautious when using these instruments. As with any instrument, counselors need to carefully evaluate any instrument before using it with a couple or family and to understand the instrument's limitations. These instruments should be viewed more as a springboard for further discussion rather than as diagnostic tools.

Summary

In recent years, hundreds of instruments have been developed for assessing couples and families. These instruments are not as well established as some of the instruments developed for use with individuals (e.g., the MMPI-2, WAIS-IV). Developing instruments for use with couples and families is more difficult than developing instruments for individuals. This chapter addressed four areas of marriage and family assessment: family interviews, mapping techniques, instruments designed for individual counseling, and instruments specific to couple and family assessment.

The family interview is probably the most frequently used assessment strategy in this area. While interviewing the family, the counselor should observe the family to note factors such as family structure and composition, family processes, relationship patterns, family affect, family organization, and family diversity. In addition, there are different mapping tasks a counselor can initiate with a family to facilitate the gathering of relevant information. One of the most commonly used techniques is the genogram. Counselors and their clients construct a family tree that records information about family members and their relationships over at least three generations. Genograms are based on the premise that family patterns can be transmitted from one generation to the next and that by recognizing these patterns a counselor can assist the family in avoiding destructive patterns.

In addition, clinicians who are counseling couples and families sometimes use instruments designed for individuals. The instruments most commonly used by marriage and family therapists are the Myers-Briggs Type Indicator, the Minnesota Multiphasic Personality Inventory-2, and the Taylor-Johnson Temperament Analysis. The T-JTA uses a system that enables the practitioner to examine congruence in the couple's personality as well as congruence between how each client views himself or herself and how each client is viewed by the spouse or significant other.

Finally, there are instruments designed specifically for use with either couples or families. The most frequently used premarital instrument is PREPARE. In addition, a number of instruments address marital satisfaction or issues after a couple is married. These instruments are typically used to identify areas in which the couple's perceptions may differ from one another. Assessment of families is probably one of the most difficult areas in counseling due to the complexities of family interactions. Some of the family assessment instruments are based on theory, and the clinician must have an understanding of the theoretical implications when using the instruments. Assessment in the area of marriage

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

and family counseling does not have the extensive history of intelligence and personality assessments. Therefore, the instruments in this area are not as well researched, nor has the field had time to adequately address many of the limitations in current assessment devices.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 15

Diagnosis

Learning Objectives

- LO1 Discuss reasons counselors should know about diagnosis
- LO2 Understand the relationship between the DSM and the ICD
- LO3 Describe the definition of a mental disorder in the *DSM-5*
- LO4 Comprehend the importance of cultural factors in diagnosis
- LO5 Distinguish between other specified and unspecified disorders
- LO6 Understand that each disorder has specific criteria
- LO7 Distinguish among the neurodevelopmental disorders
- LO8 Distinguish among the bipolar and related disorders
- LO9 Distinguish among the depressive disorders
- LO10 Distinguish among the anxiety disorders
- LO11 Distinguish among the obsessive-compulsive and related disorders
- LO12 Distinguish among the trauma- and stress-related disorders
- LO13 Distinguish among the feeding and eating disorders
- LO14 Distinguish among the elimination disorders
- LO15 Understand gender dysphoria
- LO16 Distinguish among the disruptive, impulse-control, and conduct disorders
- LO17 Distinguish among the substance-related and addictive disorders
- LO18 Distinguish among the personality disorders
- LO19 Distinguish among the paraphilic disorders
- LO20 Understand other conditions that may be the focus of clinical attention
- LO21 Identify instruments that can help with diagnosis

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

A few years ago, I worked with a woman who presented with anxiety issues particularly as they related to work. She had found that by counting by threes (i.e., 3, 6, 9, 12 . . .) she often could reduce her anxiety. If she made a mistake or could not come up with the next number soon enough she would have to start over. Sometimes she would need to count into the thousands before her anxiety would dissipate enough so that she could go to work. The reason she sought counseling is the time it took her to count by threes was increasing and she was often late for work. It was helpful to me as a counselor to recognize that she fit the symptoms of someone with obsessive-compulsive disorder. Once I correctly diagnosed her with obsessive-compulsive disorder, I could then survey the literature related to treating obsessive-compulsive disorder and select a treatment approach that was empirically supported. Therefore, having knowledge of diagnostic criteria really helped me be effective with this client because the treatment approach was very helpful and we were able to reduce her anxiety, and when she left counseling she had stopped the compulsive behavior of counting by threes.

L01 Although 25 years ago there was some debate on whether counselors should use the *Diagnostic and Statistical Manual of Mental Disorders* (Hohenshil, 1993), it is now commonly expected that in most mental health settings, counselors will need to be skilled diagnosticians. Some readers may be wondering why there is a chapter on diagnosis in a book on counseling assessment. The answer is that diagnosis is frequently part of the counseling process and the clinician must *assess* the client in order to make a diagnosis. It is not unusual for an individual to have a mental disorder as around 25% of adults suffer from a mental illness (Center for Disease Control and Prevention, 2011). Furthermore, many counselors need to know about diagnosis as a mental disorder diagnosis has generally become mandatory if medical insurance is to reimburse a clinician or an organization for treatment. Although school counselors are typically not required to provide a diagnosis, often they work with children and adolescents who have been diagnosed or they need to be familiar with diagnosis to refer a student to a certain agency that specializes in treatment for certain diagnoses. Also school counselors frequently work in multidisciplinary teams, and they need basic information about diagnostic categories.

Diagnostic systems are designed to provide a common language for professionals so that diagnostic terms have a unified meaning rather than individuals using their own unique definitions. For example, if I walked down the street and asked pedestrians for their definitions of *substance abuse,* the definitions would probably differ dramatically. Some people may not consider consumption of a six pack of beer every night as an indication of substance abuse; others may consider one drink per night to be substance abuse. For effective communication within a profession, there needs to be uniformity in definitions. Diagnostic systems attempt to provide a nosology, or nomenclature, for counselors, psychiatrists, psychologists, social workers, and other health and mental health professionals. Diagnosis should not be viewed as providing a punitive label but rather as providing a description of the client's symptoms that others can understand.

L02 In the United States, the most commonly used diagnostic system for mental disorders is the American Psychiatric Association's (APA) *Diagnostic and Statistical Manual of Mental Disorders,* which is in the fifth edition and is commonly referred to as the *DSM-5* (APA, 2013). The *DSM-5* is not the only diagnostic system for mental disorders available as much of the world uses the World Health Organization's *International Classification of Diseases* (ICD). In fact, there is a direct connection between the DSM-5 and the ICD-9-CM and the ICD-10-CM. In the *DSM-5*, both the ICD-9-CM and the ICD-10-CM codes are usually attached to relevant disorders. In the United States, clinicians will have used the ICD-9-CM

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 15.1
Previous editions of the *Diagnostic and Statistical Manual of Mental Disorders*

DSM-IV-TR (2000)
DSM-IV (1994)
DSM-III-R (1987)
DSM-III (1980)
DSM-II (1968)
DSM-I (1952)

codes through October 2015, and then after that all healthcare professionals will use the ICD-10-CM codes. The *DSM-5* lists both the ICD-9-CM and the ICD-10-CM codes; the ICD-9-CM codes are listed first and then the ICD-10-CM codes follow and are enclosed in parentheses. For example, the codes in the *DSM-5* for autism spectrum disorder are 299.00 (F84.0). Therefore, before October 2015, a clinician would report a diagnosis of autism spectrum disorder as 299.00 autism spectrum disorder; whereas after October 2015, a clinician would report F84.0 autism spectrum disorder. The organization of the *DSM-5* was also designed to correspond to the structure of the ICD-11, which is scheduled to be released in 2017.

The *DSM-5* is the latest iteration of the *Diagnostic and Statistical Manual of Mental Disorders,* and Table 15.1 includes a listing of previous editions. Rather than focusing on changes in the *DSM-5* as compared to earlier editions of the *Diagnostic and Statistical Manual of Mental Disorders,* this chapter will mainly focus on the content of the *DSM-5.* Let me start by acknowledging that I could have written an entire book describing the content of the *DSM-5.* This chapter is designed as an introduction to the *DSM-5* that will briefly overview disorders frequently seen by counselors. Each disorder within the *DSM-5* typically involves polythetic criteria, with very specific criteria that must be met in order to merit a diagnosis. This chapter will not cover the specific criteria for each disorder discussed; rather the focus is on pointing the reader to certain common disorders within the *DSM-5* and then encouraging readers to learn the specific criteria from the *DSM-5.*

The creation of the *Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition* (DSM-5) occurred over 12 years and involved hundreds of people, mostly psychiatrists, in different work groups. The *DSM-5* was designed so that there is a common nomenclature that all mental health clinicians can understand and that can aid clinicians in making reliable diagnoses. The writers of the *DSM-5* focused on the science related to the disorders and asserted that since the *DSM-IV* was published there has been, "real and durable progress in such areas as cognitive neuroscience, brain imaging, epidemiology, and genetics" (APA, 2013, p. 5), which influenced all of the working groups. Another common thread emerging from the workgroups is an understanding that the boundaries between many of the categorized disorders in the *DSM-IV* were more fluid and that some symptoms, in varying degrees of severity, may be present across many disorders. Therefore, there is a focus on a dimensional approach to mental disorders, which influenced the organization of the *DSM-5* and resulted in some cross-cutting assessments that will be discussed later. For counselors very interested in diagnosis, I would also suggest that they read Morrison (2014).

LO3 DSM-5's Definition of a Mental Disorder

It should be remembered that the *DSM-5* is primarily used to diagnose **mental disorders**, and not every difficulty a client has should be considered a mental disorder. In using the *DSM-5,* the first step is to see if the client meets the definition of a mental disorder. All disorders included in Section II of the *DSM-5* will meet the following definition

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

of a mental disorder except for "Medication-Induced Movement Disorder and Other Adverse Effects of Medication" and "Other Conditions That May Be a Focus of Clinical Attention." The following is APA's (2013) definition of a mental disorder:

> A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities. An expectable or culturally approved response to a common stressor or loss, such as the death of a loved one, is not a mental disorder. Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual, as described above. (p. 20)

It is sometimes difficult to determine if a client has a mental disorder, particularly when the presentation of symptoms is mild. A generic marker that clinicians can use is whether the disorder causes "clinically significant distress or impairment in social, occupational, or other important areas of functioning" (APA, 2013, p. 21). Once the clinician determines that a client does have a mental disorder, then he or she can turn to the diagnostic criteria for specific disorders (e.g., attention-deficit/hyperactivity disorder) to determine if the client meets the criteria for that diagnosis. The *DSM-5 makes it clear that accurate diagnosis requires training and good clinical judgment.*

Organizational Structure

As the American Psychiatric Association and the World Health Organization began to revise the *DSM* and the ICD, both focused on improving clinical utility by examining which disorders are grouped together and the overall structure of these diagnostic systems. The decision was to organize these two systems based on well-replicated science on the relationships among the diagnostic groups. For example, depression and anxiety have often been found to be co-occurring disorders so they are placed next to each other in the organizational scheme of the *DSM-5*. The goal of reorganizing similar diagnostic groups was to increase clinical utility and to encourage researchers to identify the psychological and physiological cross-cutting factors that are not bound to strict categorical designations. It may be that the organizational structure of future editions of the *DSM* may change as new evidence is found. The *DSM-5* is considered a "living document," which will be adaptable with new findings in neurobiology, genetics, and epidemiology. On the basis of current research, it was decided to organize the *DSM-5* by "internalizing and externalizing disorders" and based on developmental and life-span considerations. The *DSM-5* begins with disorders that are generally manifested early in life (e.g., neurodevelopmental and schizophrenia and other psychotic disorders). These chapters of the *DSM-5* are followed by diagnoses that are more typically manifested in adolescence and early adulthood (e.g., bipolar, depressive, and anxiety disorders). In terms of diagnoses, the *DSM-5* concludes with disorders that commonly occur in adulthood or later (i.e., neurocognitive disorders). This organization structure is designed to improve clinical utility by encouraging clinicians to strongly consider the client's developmental level. The same organizational structure has been taken whenever possible with chapters of the *DSM-5*, where those disorders usually occurring in childhood are provided first, followed by disorders occurring in adolescence and adulthood, and finally with disorders that are typically manifested in late adulthood. Therefore, the overall organizational structure of the *DSM-5* is as follows: neurodevelopmental disorders, internalizing disorders, externalizing disorders, neurocognitive disorders, and other disorders. Table 15.2 provides a listing of the chapters included in Section II of the *DSM-5*.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 15.2
Section II: Chapter Structure

- Neurodevelopmental Disorders
- Schizophrenia Spectrum and Other Psychotic Disorders
- Bipolar and Related Disorders
- Depressive Disorders
- Anxiety Disorders
- Obsessive-Compulsive and Related Disorders
- Trauma- and Stress-Related Disorders
- Dissociative Disorders
- Somatic Symptom and Related Disorders
- Feeding and Eating Disorders
- Elimination Disorders
- Sleep-Wake Disorders
- Sexual Dysfunctions
- Gender Dysphoria
- Disruptive, Impulse-Control, and Conduct Disorders
- Substance-Related and Addictive Disorders
- Neurocognitive Disorders
- Personality Disorders
- Paraphilic Disorders
- Other Mental Disorders
- Medication-Induced Movement Disorders and Other Adverse Effects of Medication
- Other Condition That May Be a Focus of Clinical Attention (V-codes and Z-codes)

LO4 Cultural Issues

According to the APA (2013), "mental disorders are defined in relation to culture, social, and familial norms and values" (p. 14). Hence, clinicians need to be sensitive to cultural issues when diagnosing clients. In Section III of the *DSM-5* under Cultural Formulation, there is a discussion of culture and diagnosis, which includes the Cultural Formulation Interview (CFI). The CFI contains 16 questions designed to assist clinicians in obtaining pertinent information regarding culture. The CFI is best used with demographic information so the questions can be tailored to the specific client being assessed. Also in the Appendix of the *DSM-5* is a "Glossary of Cultural Concepts of Distress," which describes common cultural syndromes, idioms of distress, and causal explanation that are relevant to clinical practice.

Nonaxial System

With the *DSM-5*, the counselor no longer has to provide a five-axial diagnosis. If you are not familiar with a five-axial diagnosis, then it makes it easier to learn the *DSM-5* system. Previously different information was listed on Axes I, II, III, IV, and V (e.g., medical conditions were listed on Axis III). Medical information, psychosocial, and contextual factors continue to be reported; it just does not involve an axial system. I will use the ICD-10-CM and the *DSM-5* to demonstrate a diagnosis with a nonaxial system.

F34.0 Cyclothymic disorder
J30.9 Rhinitis, allergic
Z59.6 Low income

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

As we progress in our learning about the *DSM-5*, these codes, what they represent, and where they were found will become more evident. The first code, F34.0 Cyclothymic disorder, comes from Section II of the *DSM-5*, and it is a mental disorder. J30.9 Rhinitis, allergic involves physical symptoms related to allergies and came from the ICD-10-CM. Finally, Z59.6 Low income also comes from the *DSM-5*, and this code comes from the section of the *DSM-5* labeled "Other Conditions That May Be a Focus of Clinical Attention." The section "Other Conditions That May Be a Focus of Clinical Attention" is also known as the V-codes or Z-codes.

It is possible to have more than one diagnosis of a mental disorder, and they are usually listed in terms of ascending order of importance. In an inpatient setting, the primary diagnosis is listed first and a phrase (principal diagnosis) typically follows. In an outpatient setting as well, the primary diagnosis is listed first and the phrase (reason for visit) follows. It is also possible to give a provisional diagnosis (e.g., when the clinician suspects there may be a personality disorder but is unsure). In this case, the clinician lists the code number and the diagnosis, followed by a specifier (provisional).

LO5 Other Specified and Unspecified Disorders

When diagnosing a client, counselors have the option of using the *other specified disorder* or the *unspecified disorder* if a client does not meet the full criteria necessary for a specific diagnosis. The other specified disorder allows the clinician to communicate the specific reasons that the client's presentation does not meet the criteria for a specific disorder. For example, if a client has been experiencing a depressed mood for 10 days but not for the required 2 weeks, then the clinician could use the other specified depressive disorder label and comment that the client did not reach the two-week requirement. If the clinician chooses not to specify which criteria are not met for a disorder, then he or she can use the unspecified disorder option. Note that the other specified or unspecified disorders are based on clinical judgment and provide flexibility in diagnosing.

Subtypes and Specifiers

Whereas other specified and unspecified disorders are used when a client's symptoms do not meet the full criteria for a diagnosis, subtypes and specifiers are used to more clearly delineate the client's symptoms within a diagnosis. Therefore, a clinician uses subtypes or specifiers after the clinician has determined that the full criteria for a diagnosis are met. *Subtypes* define mutually exclusive and jointly exhaustive subgroupings within a diagnosis. For example, within the diagnosis of attention-deficit/hyperactivity disorder (ADHD), there are three subtypes: predominantly inattentive presentation, predominantly hyperactive/impulsive presentation, and combined presentation. If you need to use a subtype, it will be indicated by the term *specify whether. Specifiers* are not intended to be mutually exclusive, and, therefore, you can use more than one. Specifiers simply provide more information and often involve specifying whether the disorder is mild, moderate, or severe. If you are confused about specifiers, refer to a diagnosis within the *DSM-5*. I would suggest that you look at major depressive disorder, and you can see that you specify by severity (mild, moderate, severe, etc.) and by some features (with anxious distress, with mixed features, with melancholic features, etc.). There are descriptions for the specifiers so that the clinician is guided in specifying or providing more detail to the diagnosis.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

DSM-5 Emerging Measures

In developing the *DSM-5*, the workgroups also developed emerging measures to assist clinicians in the diagnostic process. This work is continuing and new instruments continue to be developed. Some of the instruments are included in Section III of the *DSM-5*, and other instruments can be examined at www.dsm5.org. Included in Section III of the *DSM-5* is the Level 1 Cross-Cutting Symptom Measure. There are two forms of the Level 1 Cross-Cutting Symptom Measure: an adult version and a parent/guardian version that is designed for children ages 6 to 17. Figure 15.1 shows the DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure—Adults. Level 1 cross-cutting measures were developed to be administered at the initial patient interview and to monitor treatment progress.

Figure 15.1
DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure—Adults. ***http://www.psychiatry.org/practice/dsm/dsm5/online-assessment-measures#Level1 American Psychiatric Association***

DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure–Adult

Name: _____ Age: ____ Sex: ❑ Male ❑ Female Date_____

If this questionnaire is completed by an informant, what is your relationship with the individual? _____
In a typical week, approximately how much time do you spend with the individual? _____ hours/week

Instructions: The questions below ask about things that might have bothered you. For each question, circle the number that best describes how much (or how often) you have been bothered by each problem during the past TWO (2) WEEKS.

	During the past TWO (2) WEEKS, how much (or how often) have you been bothered by the following problems?	None Not at all	Slight Rare, less than a day or two	Mild Several days	Moderate More than half the days	Severe Nearly every day	Highest Domain Score (clinician)
I.	1. Little Interest or pleasure in doing things?	0	1	2	3	4	
	2. Feeling down, depressed, or hopeless?	0	1	2	3	4	
II.	3. Feeling more irritated, grouchy, or angry than usual?	0	1	2	3	4	
III.	4. Sleeping less than usual, but still have a lot of energy?	0	1	2	3	4	
	5. Starting lots more projects than usual or doing more risky things than usual?	0	1	2	3	4	
IV.	6. Feeling nervous, anxious, frightened, worried, or on edge?	0	1	2	3	4	
	7. Feeling panic or being frightened?	0	1	2	3	4	
V.	8. Avoiding situations that make you anxious?	0	1	2	3	4	
	9. Unexplained aches and pains (e.g., head, back, joints, abdomen, legs)?	0	1	2	3	4	
	10. Feeling that your illnesses are not being taken seriously enough?	0	1	2	3	4	
VI.	11. Thoughts of actually hurting yourself?	0	1	2	3	4	
VII.	12. Hearing things other people couldn't hear, such as voices even when no one was around?	0	1	2	3	4	
	13. Feeling that someone could hear your thoughts, or that you could hear what another person was thinking?	0	1	2	3	4	
VIII.	14. Problems with sleep that affected your sleep quality over all?	0	1	2	3	4	
IX.	15. Problems with memory (e.g., learning new information) or with location (e.g., finding your way home)?	0	1	2	3	4	
X.	16. Unpleasant thoughts, urges, or images that repeatedly enter your mind?	0	1	2	3	4	
	17. Feeling driven to perform certain behaviors or mental acts over and over again?	0	1	2	3	4	
XI.	18. Feeling detached or distant from yourself, your body, your physical surroundings, or your memories?	0	1	2	3	4	
XII.	19. Not knowing who you really are or what you want out of life?	0	1	2	3	4	
	20. Not feeling close to other people or enjoying your relationships with them?	0	1	2	3	4	
XIII.	21. Drinking at least 4 drinks of any kind of alcohol in a single day?	0	1	2	3	4	
	22. Smoking any cigarettes, a cigar, or pipe, or using snuff or chewing tobacco?	0	1	2	3	4	
	23. Using any of the following medicines ON YOUR OWN, that is, without a doctor's prescription, in greater amounts or longer than prescribed [e.g., painkillers (like Vicodin), stimulants (like Ritalin or Adderall), sedatives or tranquilizers (like sleeping pills or Valium), or drugs like marijuana, cocaine or crack, club drugs (like ecstasy), hallucinogens (like LSD), heroin, inhalants or solvents (like glue), or methamphetamine (like speed)]?	0	1	2	3	4	

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

The instructions clearly indicate that these measures were designed to enhance clinical decision making and not as the sole basis for making a clinical diagnosis. The DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure—Adults briefly assesses the areas of depression, anger, mania, anxiety, somatic symptoms, suicidal ideation, psychosis, sleep problems, memory, repetitive thoughts and behaviors, dissociation, personality functioning, and substance use. For many of these domains there are Level 2 measures, which address the specific domain. The Level 2 measures can also be downloaded and used from the website (i.e., www.dsm5.org).

Readers familiar with the *DSM-IV-TR* will be aware of the Global Assessment of Functioning (GAF) scores. Clinicians no longer need to report GAF scores in the *DSM-5*. If clinicians want a measure that resembles a measure of overall functioning, they can use the World Health Organization Disability Schedule (WHODAS 2.0), which is a measure of disability. The WHODAS 2.0 is a 36-item instrument that measures disability across six domains: understanding and communicating, getting around, self-care, getting along with people, life activities, and participation in society. Clinicians can receive a total score or scores across the six domains. Information on the WHODAS 2.0 is included in Section III of the *DSM-5* (APA, 2013).

L06 Diagnostic Criteria

Mental disorders are categorically classified in the *DSM-5* based on the practitioner's evaluation of certain criteria. The *DSM-5* covers a wide array of diagnostic categories, with applications for clinicians in educational, clinical, and research settings. The *DSM-5* clearly stipulates the criteria that need to be present to meet the diagnostic conditions. For example, the diagnostic criteria for *conduct disorder* are as follows:

A. A repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rules are violated, as manifested by the presence of at least three of the following 15 criteria in the past 12 months from any criteria below, with at least one criterion present in the past 6 months:

Aggression to people and animals

(1) often bullies, threatens, or intimidates others.
(2) often initiates physical fights.
(3) has used a weapon that can cause serious physical harm to others (e.g., a bat, brick, broken bottle, knife, gun).
(4) has been physically cruel to people.
(5) has been physically cruel to animals.
(6) has stolen while confronting a victim (e.g., mugging, purse snatching, extortion, armed robbery).
(7) has forced someone into sexual activity.

Destruction of property

(8) has deliberately engaged in fire setting with the intention of causing serious damage.
(9) has deliberately destroyed others' property (other than by fire setting).

Deceitfulness or theft

(10) has broken into someone else's house, building, or car.
(11) often lies to obtain goods or favors or to avoid obligations (i.e., "cons" others).
(12) has stolen items of nontrivial value without confronting a victim (e.g., shoplifting, but without breaking and entering; forgery).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Serious violations of rules

(13) often stays out at night despite parental prohibitions, beginning before age 13 years

(14) has run away from home overnight at least twice while living in parental or parental surrogate home or once without returning for a lengthy period

(15) is often truant from school, beginning before age 13 years

B. The disturbance in behavior causes clinically significant impairment in social, academic, or occupational functioning.

C. If the individual is age 18 years or older, criteria are not met for antisocial personality disorder.

Specify whether:

312.81 (F91.1) Childhood-onset type: Individuals show at least one symptom characteristic of conduct disorder prior to age 10.

312.82 (F91.2) Adolescent-onset type: Individuals show no symptoms characteristic of conduct disorder prior to age 10.

312.89 (F91.3) Unspecified onset: Criteria for a diagnosis of conduct disorder are met, but there is not enough information available to determine whether the onset of the first symptom was before or after the age of 10.

Specify if:

With limited prosocial emotions: To qualify for this specifier, an individual must have displayed at least two of the following characteristics persistently over 12 months and in multiple relationships and settings. These characteristics reflect the individual's typical pattern of interpersonal and emotional functioning over this period and not just occasional occurrences in some situations. Thus, to assess the criteria for the specifier, multiple information sources are necessary. In addition to the individual's report, it is necessary to consider reports by others who have known the individual for extended periods of time (e.g., parents, teachers, co-workers, extended family members, peers).

Lack of remorse or guilt: Does not feel bad or guilty when he or she does something wrong (exclude remorse when expressed only when caught or facing punishment). The individual shows a general lack of concern about the negative consequences of his or her actions. For example, the individual is not remorseful after hurting someone or does not care about the consequences of breaking rules.

Callous—lack of empathy: Disregards and is unconcerned about the feelings of others. The individual is described as cold and uncaring. The person appears more concerned about the effects of his or her actions on himself or herself, rather than their effects on others, even when they result in substantial harm to others.

Unconcerned about performance: Does not show concern about poor/problematic performance at school, at work, or in other important activities. The individual does not put forth the effort necessary to perform well, even when expectations are clear, and typically blames others for his or her poor performance.

Shallow or deficient affect: Does not express feelings or show emotions to others, except in ways that are shallow, insincere, or superficial (e.g., actions contradict the emotion displayed, can turn emotions "on" or "off" quickly) or when emotional expressions are used for gain (e.g., emotions displayed to manipulate or intimidate others).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Specify current severity:

Mild: Few if any conduct problems in excess of those required to make the diagnosis are present and conduct problems cause relatively minor harm to others (e.g., lying, truancy, staying out after dark without permission, other rule breaking).

Moderate: The number of conduct problems and the effect on others are intermediate between those specified in "mild" and those in "severe" (e.g., stealing without confronting a victim, vandalism).

Severe: Many conduct problems in excess of those required to make the diagnosis are present, or conduct problems cause considerable harm to others (e.g., forced sex, physical cruelty, use of a weapon, stealing while confronting a victim, breaking and entering). (APA, 2013, pp. 469–471)

Common Diagnoses

Although I will not cover every diagnosis that is included in Section II of the *DSM-5*, I will try to briefly describe the major characteristics of many of the common diagnoses that counselors should be familiar. Moreover, I will not be discussing the other specified or unspecified disorders, which were discussed earlier, but readers should remember these options. As indicated earlier, these will be brief overviews of the diagnoses that are designed to provide readers with a general understanding of the disorder, which they can then use to examine the specific criteria in the *DSM-5*. There are 22 chapters with diagnoses (see Table 15.2), which correspond to the organizational structure mentioned earlier, and I will be discussing them in order.

L07 Neurodevelopmental Disorders

The neurodevelopmental disorders typically manifest during the developmental period and are characterized by developmental deficits that result in impairments of personal, social, academic, or occupational functioning. The neurodevelopmental disorders often co-occur. For example, many children with ADHD also have specific learning disorder (APA, 2013). *Intellectual disability (intellectual developmental disorder)* was called mental retardation and is characterized by deficit in intellectual functioning as confirmed by both clinical assessment and individualized, standardized intelligence testing. In addition, there needs to be deficit in adaptive functioning, which was described in Chapter 6. Furthermore, the deficits in intellectual and adaptive functioning must be present during the developmental period. *Global developmental delay* is reserved for children under the age of 5 who have failed to meet developmental milestones in several areas of intellectual functioning. It includes individuals who are too young to participate in standardized testing, and this category requires future reassessment after a time period.

In the chapter on neurodevelopmental disorders, there is a subcategory label of communication disorders. Counselors do not typically diagnose these disorders, which are *language disorder, speech sound disorder, childhood-onset fluency disorder (stuttering),* and *social (pragmatic) communication disorder.*

Autism spectrum disorder has been changed from the *DSM-IV-TR* and encompasses the previous diagnoses of Asperger's disorder, childhood disintegrative disorder, and pervasive developmental disorder not otherwise specified. Autism spectrum disorder is characterized by symptoms in two core areas: (a) deficits in social communication and social interactions, and (b) restricted repetitive behaviors, interests, and activities. Symptoms must be present in the early developmental period, but they may not be fully manifested until later in life.

Attention-deficit/hyperactivity disorder (ADHD) may be a disorder many readers are familiar with as there are some questions related to over-diagnosis. Children must have six or more symptoms in either the inattention domain or the hyperactivity/impulsivity

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

domain or six in each for the *combined presentation.* Adults must have at least five symptoms. There are three subtypes of ADHD: *combined presentation, predominantly inattentive presentation,* and *predominantly hyperactive/impulse presentation.* In addition, several symptoms of either inattention or hyperactivity-impulsivity must be present before the age of 12 and in two or more settings (e.g., home, school, or work).

Specific learning disorder involves difficulties with learning and specific academic skills despite the provision of interventions that target those difficulties. Therefore this criterion is consistent with the response-to-intervention approach discussed in Chapter 10. The affected academic skills must be substantially and quantifiably below those expected for the individual's chronological age and cause significant interference in academic or occupational functioning or with daily living. There are three disorders: *specific learning disorders with impairment in reading, specific learning disorders with impairment in written expression,* and *specific learning disorders with impairment in mathematics.* These disorders are confirmed by individually administered standardized measures and comprehensive clinical assessment. For clients age 17 and older, a documented history of impaired learning difficulties can be substituted for the standardized assessment. The learning difficulties begin during the school-aged years but may not become fully manifested until the demands for those academic skills exceed the individual's capabilities (e.g., timed tests, complex writing assignments, or a heavy academic load).

There are also *motor disorders,* which is included in the chapter of neurodevelopmental disorders. Readers interested in either *developmental coordination disorder* and *stereotypic movement disorder* should see the *DSM-5. Tic disorders* are also included in this chapter. A tic is "a sudden, rapid, recurrent, nonrhythmic motor movement or vocalization" (APA, 2013, p. 81). For *Tourette's disorder,* the individual must have both multiple motor and one of more vocal tics. In addition, the tics may wax and wane in frequency but they must be present for over a year. Clients can also have *persistent (chronic) motor or vocal tic disorder.* If the tics are not present for a year, then the clinician can use *provisional tic disorder.*

Schizophrenia Spectrum and Other Psychotic Disorders

All of the disorders in this section are characterized by having psychotic symptoms as the defining feature. Because counselors are primarily trained with a developmental approach, this area will be only briefly summarized. *Delusional disorder* involves the presence of one or more delusions that last for a month or longer. Hallucinations, if present, are not prominent and are related to the delusions. There are subtypes of delusional disorder. *Brief psychotic disorder* is manifested by having at least one of the symptoms of delusions, hallucinations, disorganized speech, or grossly disorganized or catatonic behaviors and one of the symptoms must be either delusions, hallucinations, or disorganized speech. The duration of the episode is at least one day and less than a month. *Schizophreniform disorder* has at least two of the above symptoms (i.e., delusions, hallucinations, disorganized speech, etc.) with at least one being delusions, hallucinations, or disorganized speech. The duration is from one to six months. *Schizophrenia* varies from schizophreniform disorder in the duration must be at least six months, but it still has the criteria of at least two of the above symptoms, with at least one being delusions, hallucinations, or disorganized speech. *Schizoaffective disorder* involves an uninterrupted period involving either a major depressive or a manic episode and the symptoms of schizophrenia. At some point during the illness, delusions or hallucinations have been present for two or more weeks in the absence of a major depressive or manic episode; however, during the majority of time, the symptoms of either a major depressive or a manic episode are present. Also in this chapter are the diagnoses of *substance/medication-induced psychotic disorder, psychotic disorder due to another medical condition,* and *catatonia associated with another mental disorder.*

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L08 Bipolar and Related Disorders

Before counselors can understand the criteria for diagnosing either bipolar disorders or depressive disorders, they need to understand *mood episodes* (i.e., manic episode, hypomanic episode, and major depressive episode). Mood episodes are *not* diagnosed as disorders; instead, they are the building blocks for diagnoses. A **manic episode** is characterized by a distinct period of abnormally elevated, expansive, or irritable moods and abnormally increased goal-directed activity or energy that lasts at least *one week* and is present most of the day. During this period, the individual exhibits three (four if the mood is irritable) of the following symptoms and these symptoms represent a notable change: (1) inflated self-esteem or grandiosity, (2) decreased need for sleep, (3) more talkative than usual or pressure speech, (4) flight of ideas or subjective experience that thoughts are racing, (5) distractibility, (6) increase in goal-directed activity (socially, at work, or sexually) or psychomotor agitation (i.e., purposeless non-goal-directed activity), (7) excessive involvement in activities that have a high potential of painful consequence (e.g., excessive purchasing, sexual indiscretions, foolish business investments) (APA, 2013). A manic episode is not a normal feeling good about oneself; rather, it is a feeling of becoming so driven that it causes marked impairments in occupational functioning, social activities, or relationships.

A **hypomanic episode** is characterized by the same symptoms as a manic episode except that it lasts for only *four days.* In addition, the episode is not severe enough to cause marked impairment in occupational or social function and does not require hospitalization. If there are psychotic features, then it is considered a manic episode. The symptoms of a hypomanic episode are, however, an unequivocal change in functioning and are noticeable by others.

A **major depressive episode** is one in which the client is in a depressed mood or has lost interest or pleasure in nearly all activities for at least two weeks. In children and adolescents, the mood may be one more of irritability than of sadness. In addition, the client must experience five or more of the following symptoms, with at least one of the symptoms being either depressed mood or loss of interest or pleasure: (1) depressed mood most of the day, nearly every day (note with children and adolescents, the mood can be irritable); (2) markedly diminished interest or pleasure in all, or most all, activities most of the day, nearly every day; (3) significant changes in appetite or weight; (4) changes in sleep (insomnia or hypersomnia); (5) psychomotor agitation or retardation (observed by others); (6) fatigue or loss of energy; (7) feelings of worthlessness or guilt; (8) difficulty thinking, concentrating, or making decisions; or (9) recurrent thoughts of death or suicidal ideation, suicide plans, or attempts. These symptoms need to persist through most of the day and cannot be the direct result of a medical condition.

For a diagnosis of *bipolar I,* the criteria is that the client must have had at least one manic episode. Although some clients will also have a major depressive episode, this is not required for a diagnosis of bipolar I. Interested readers are directed to the *DSM-5* on how to code for bipolar I and how to specify features that may be associated with the disorder. The diagnostic criteria for *bipolar II* are that the client either currently or in the past has suffered a hypomanic episode and a major depressive episode. There are also coding procedures for a diagnosis of bipolar II. Somewhat different from bipolar I and bipolar II is the *cyclothymic disorder,* in which a client needs to have had numerous periods of hypomanic symptoms that do not meet the criteria for a manic episode and numerous periods of depressive symptoms that do not meet the criteria for a major depressive episode. These symptoms must last at least two years for adults and one year for children and adolescents. In this chapter, there are also diagnoses of *substance/medication-induced bipolar and related disorder* and *bipolar and related disorder due to a medical condition.*

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

L09 Depressive Disorders

A diagnosis that is new to the *DSM-5* is *disruptive mood dysregulation disorder.* This disorder is characterized by severe and recurrent temper outburst, which can be either verbal or behavioral, and are grossly out of proportion in intensity or duration to the situation or provocation. These temper outbursts are not consistent with the developmental level and occur, on average, three or more times per week. Furthermore, the mood between the outbursts is persistently irritable or angry. This diagnosis should not be made before the age of 6 nor after the age of 18.

A *major depressive disorder* has the same criteria as a major depressive episode (see the earlier discussion). In previous versions of the *DSM*, grief or bereavement was excluded from the diagnosis of major depressive disorder; however, with the advent of the *DSM-5*, it is now possible to diagnose someone suffering from grief or bereavement. There is a note in this diagnosis that says, "responses to a significant loss (e.g., bereavement, financial ruin, losses from a natural disaster, a serious medical illness, or disability) may include the feelings of intense sadness, rumination about the loss, insomnia, poor appetite, and weight loss noted in Criterion A, which may resemble a depressive episode. Although such symptoms may be understandable or considered appropriate to the loss, the presence of a major depressive episode in addition to the normal response to a significant loss should also be carefully considered. This decision inevitably requires the exercise of clinical judgment based on the individual's history and the cultural norms for the expression of distress in the context of loss" (APA, 2013, p. 161). Clinicians should also consult the *DSM-5* for clarifications on coding for major depressive disorder.

Another diagnosis in this chapter is *persistent depressive disorder (dysthymia).* This diagnosis is appropriate when a client has a depressed mood for most of the day, for more days than not for at least two years. For children and adolescents, the mood can be irritable and the duration must be for at least a year. In addition, the client must have two or more of the following symptoms: poor appetite or overeating, insomnia or hypersomnia, low energy or fatigue, low self-esteem, poor concentration or difficulty making decisions, and feelings of hopelessness (APA, 2013).

A somewhat controversial diagnosis is *premenstrual dysphoric disorder.* With premenstrual dysphoric disorder, the client must have five symptoms (one from the first group) in the final week before the onset of menses, which start to improve within a few days after the onset of menses and become minimal or absent in the week postmenses. One of the following symptoms must be present, which counts toward the five necessary symptoms: marked affective lability; marked irritability or anger or increased interpersonal conflicts; marked depressed mood, feeling hopeless, or self-deprecating thoughts; or marked anxiety, tension, and/or feelings of being keyed up or on edge. In addition, one or more of the following symptoms must be present to reach a total of five combined symptoms: decreased interest in typical activities; difficulties in concentration; lethargy, fatigue, or marked loss of energy; marked change in appetite or specific food cravings, a sense of being overwhelmed or out of control; or physical symptoms such as breast tenderness. Clinicians should examine the *DSM-5* for methods for confirming this diagnosis with daily ratings. Also in the chapter on depressive disorders are *substance/medication-induced depressive disorder* and *depressive disorder due to another medical condition.*

L10 Anxiety Disorders

As you will remember, the chapters of the *DSM-5* are arranged in developmental order so that disorders associated with childhood are toward the beginning of the chapter, which is true with anxiety disorders. The first diagnosis listed in this chapter is *separation anxiety*

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

disorder, which frequently occurs in early childhood although it can occur later. Separation anxiety disorder is developmentally inappropriate fear and anxiety concerning separation from someone whom the client is attached. The fear and anxiety must last at least four weeks in children and adolescents and six months in adults. The second disorder in this chapter is *selective mutism,* which involves consistent failure to speak in specific social situations in which speech is expected (e.g., school) despite speaking in other situations.

The third disorder in this chapter is *specific phobia.* The first criterion is marked fear or anxiety about a specific situation (e.g., flying, seeing blood) or object (e.g., spiders). Exposure to the phobic object or situation almost always provokes immediate fear or anxiety. The phobic situation or object is actively avoided or endured with intense fear or anxiety. Furthermore, the fear or anxiety is out of proportion to the actual situation and sociocultural norms. In addition, this problem is persistent and usually last at least six months. Clinicians should also see the *DSM-5* for coding.

Social anxiety disorder (social phobia) is a similar disorder to specific phobia, except that marked fear and anxiety is to one or more social activities where the client could be exposed to possible scrutiny by others. In children, the fear and anxiety must occur in peer settings and not solely with adults. The client fears that he or she will show his or her anxiety and will be negatively evaluated by others (e.g., embarrassed). Social situations almost always provoke anxiety and are avoided or endured with intense fear or anxiety.

Panic disorder involves the recurrence of *panic attacks.* A **panic attack** is an abrupt surge of fear or intense discomfort that reaches its peak quickly, and clients experience symptoms such as palpitations, sweating, trembling, choking, chest pain, nausea, and fear of losing control. Furthermore, individuals must have either persistent worry or concern about having another panic attack or significant maladaptive behaviors related to the panic attacks. This persistent worry or change in behavior must last for a month or more.

Agoraphobia is another anxiety disorder, and it involves intense fear or anxiety about two or more of the following: (1) using public transportation, (2) being in open spaces, (3) being in enclosed spaces, (4) standing in a line or being in a crowd, and (5) being outside of the home alone. The agoraphobic situations almost always provoke fear and anxiety, and the person fears or avoids these situations. The fear, anxiety, or avoidance is persistent and lasts for at least six months.

Another important diagnosis for counselors to know is *generalized anxiety disorder.* Generalized anxiety disorder is somewhat different from the other anxiety disorders because it involves more generalized anxiety and worry that tends to persist for six months or more. An individual with this disorder finds it difficult to control the worry. In addition, the anxiety and worry result in symptoms such as restlessness, feeling keyed up, being easily fatigued, difficulty concentrating, irritability, muscle tension, and sleep disturbance. Also in the anxiety disorder section of the *DSM-5* are diagnoses of *substance/medication-induced anxiety disorder* and *anxiety due to another medical condition.*

L11 Obsessive-Compulsive and Related Disorders

Obsessive-compulsive disorder is characterized by recurrent obsessions or compulsions or both severe enough to be time consuming or to cause marked distress or impairment. Obsessions are persistent thoughts, urges, or images that are experienced as intrusive and unwanted, and in most individuals cause anxiety and fear. Furthermore, the individual attempts to neutralize these thoughts, urges, or images. Compulsions are repetitive behaviors (e.g., checking the stove, hand-washing) or mental acts (e.g., counting, repeating certain words) that the individual feels driven to perform. These behaviors or mental acts are aimed at reducing the anxiety; however, these behaviors or mental acts are clearly excessive and not connected in a realistic manner to what they are designed to neutralize or

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

prevent. Simply checking the stove a couple of times will not meet the criteria for obsessive-compulsive disorder because the obsessions or compulsions need to be problematic and/or time consuming (at least an hour a day).

Another obsessive-compulsive disorder is *body dysmorphic disorder* in which the person is preoccupied with one or more perceived defect or flaw in physical appearance that appears slight or nonobservable to others. In response to these appearance concerns, the individual performs repetitive behaviors, such as mirror checking, skin picking, or reassurance seeking. A new diagnosis to the *DSM-5* is *hoarding disorder,* where the individual has persistent difficulty discarding or parting with possessions regardless of their value. This difficulty is due to the perceived need to save the items and the distress associated with discarding them. The difficulty with discarding items results in the accumulation of possessions that is problematic. *Trichotillomania (hair-pulling disorder)* involves recurrent pulling of one's hair, which results in hair loss. The client attempts to decrease or stop the hair pulling but is unable to stop. *Excoriation (skin-picking) disorder* concerns recurrent skin picking that leads to skin lesions. Once again, the client attempts to decrease or stop the skin picking but is unable to stop. In this chapter, there is also *substance/medication-induced obsessive-compulsive and related disorder* and *obsessive-compulsive disorder due to another medical condition.*

L12 Trauma- and Stress-Related Disorders

Counselors rarely diagnose *reactive attachment disorder,* but it is a disorder characterized by inhibited, emotionally withdrawn behavior and a persistent social and emotional disturbance that are a result of a child experiencing a pattern of extreme insufficient care. *Disinhibited social engagement disorder* is also the result of extreme insufficiency in care; however, it is characterized by a pattern in which the child actively approaches and interacts with unfamiliar adults and exhibits socially disinhibited behavior.

Another important diagnosis for counselor to be aware of is *posttraumatic stress disorder* (PTSD). For counselors working with children under the age of 6, there are specific criteria in the *DSM-5*, which I will not address. With posttraumatic stress disorder, the person must first experience a traumatic event. They then must experience intrusive symptoms associated with the traumatic event (e.g., distressing memories, distressing dreams, dissociative reactions, prolonged psychological distress, marked physical reactions). There also needs to be persistent avoidance of stimuli associated with the traumatic event(s). Furthermore, there needs to be negative alterations in cognition and mood associated with the traumatic events(s). Clients with PTSD will also have marked changes in arousal and reactivity beginning or becoming worse after the traumatic event(s). In addition, the posttraumatic symptoms are present not only immediately following the traumatic event but continue for more than a month.

Acute stress disorder has many of the same symptoms as PTSD, such as experiencing a traumatic event. Also, the individual needs to have symptoms related to intrusion, negative mood, dissociative symptoms, avoidance symptoms, and arousal symptoms. In terms of duration, the symptoms should last at least three days and less than a month.

Adjustment disorder is a common diagnosis that counselors should be familiar with. The individual develops emotional or behavioral symptoms in response to an identifiable stressor(s) (e.g., divorce). The symptoms are manifested within three months of the onset of the stressor and do not last more than six months. The reaction to the stressor(s) must be either marked distress or significant impairment in social, occupational, or other important areas of functioning. Clinicians should consult the *DSM-5* on how to code for adjustment disorders.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Dissociative Disorders

Dissociative disorders are "characterized by a disruption of and/or discontinuity in normal integration of consciousness, memory, identity, emotion, perception, body representation, motor control, and behavior" (APA, 2013, p. 291). In my experience, these disorders are typically not diagnosed by counselors. *Dissociative identity disorder* (formerly multiple personality disorder) involves the presence of two or more distinct identities or personality states that recurrently take control of the client's behavior. These distinct identities or personalities are accompanied by the inability to recall important personal information that is too extensive to be explained by ordinary forgetfulness. *Dissociative amnesia* is characterized by the client's inability to recall important personal information. The information is usually of a traumatic or stressful nature and is too extensive to be explained by ordinary forgetfulness. *Depersonalization/derealization disorder* consists of the presence of persistent or recurrent experiences of depersonalization, derealization, or both. Depersonalization involves the experience of being detached or being an outside observer of one's thoughts, feelings, sensations, or actions. Derealization, on the other hand, involves experiences of unreality or detachment with respect to one's surroundings (e.g., individuals or objects are experiences as foggy, dreamlike, or distorted). Even though the person is experiencing the depersonalization or derealization, his or her reality testing remains intact.

Somatic Symptom and Related Disorders

With somatic symptom and related disorders the focus is on bodily symptoms. *Somatic symptom disorder,* formerly known as somatization disorder, there is one or more somatic symptom that results in significant disruption in daily functioning and there is an excessive focus on those symptoms. *Illness anxiety disorder,* formerly known as hypochondriasis, concerns a preoccupation with having a serious illness. *Conversion disorder (functional neurological disorder)* is where the client exhibits neurological symptoms, but the clinical findings show an incompatibility between the symptoms and recognized neurological or medical conditions. *Psychological factors affecting other mental condition* is a diagnosis when a client has a medical condition and psychological or behavioral factors that adversely affect the medical condition. *Factitious disorder* is when the client falsely presents himself or herself as ill, impaired, or injured without any external rewards, but rather because they like the role of the sick person.

L13 Feeding and Eating Disorders

Feeding and eating disorders involve disturbances in eating behaviors. *Pica* involves the eating of nonnutritive, nonfood substances, which persists for at least a month. *Rumination disorder* concerns the repeated regurgitation (rechewing) of food for over a month. A new diagnosis to the *DSM-5* is *avoidant/restrictive food intake disorder,* which is when the client fails to eat enough to meet nutritional or energy needs. An example of this is when a child fails to grow because he or she rarely eats.

Another important diagnosis for counselors to be aware of is *anorexia nervosa.* The central feature of anorexia nervosa is the inability to maintain a minimal normal body weight. The client, even though underweight, is intensely afraid of gaining weight and exhibits a significant disturbance in the perception of his or her body. Some individuals feel globally overweight; others are concerned with certain parts of their body being "too fat." The mean age of onset for this disorder is 17, although some studies indicate that there are peaks around the age of 14 and around the age of 18. There are two subtypes, one in which the food intake is severely restricted and the other of the binge eating/purging type.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Bulimia nervosa is an eating disorder characterized by binge eating and inappropriate compensatory behaviors to prevent weight gain. Clients with bulimia nervosa have recurrent episodes of binge eating during which, within 2 hours, they eat large amounts of food. These binges do not involve slight overeating; the amount of food is large. Clients report a sense of lack of control over eating during these binge episodes. To prevent gaining weight, the person uses recurrent inappropriate compensatory behaviors (e.g., self-induced vomiting, use of laxatives or diuretics, or excessive exercise). The bulimia nervosa diagnosis is given when the binging and purging occur at least once a week for a three-month period. Similar to clients with anorexia nervosa, clients with bulimia nervosa have self-images that are unduly influenced by their body shape and weight. Clients with bulimia nervosa, however, are much more realistic in their perceptions of their own bodies as compared with anorexic clients.

A somewhat controversial diagnosis is *binge-eating disorder.* Similar to bulimia nervosa, clients have recurrent episodes of binge eating during which, within 2 hours, they eat large amounts of food. The binge eating involves eating more rapidly than normal; eating until feeling uncomfortable; eating large amounts of food when they are not hungry; embarrassment about eating; and feeling depressed, guilty, or disgusted after eating. The binge eating needs to occur at least once a week for three months, and clients need to feel distressed about their eating.

L14 Elimination Disorders

Elimination disorders include *enuresis,* the repeated voiding of urine into inappropriate places, and *encopresis,* the repeated passage of feces into inappropriate places. For enuresis the voiding of urine needs to occur twice a week for three months; whereas, for encopresis the passage of feces needs to occur once a week for three months. Enuresis cannot be diagnosed before the age of 4, while for encopresis the age is 5.

Sleep-Wake Disorders

Counselors are not typically involved in the diagnosis of sleep-wake disorders (e.g., narcolepsy) unless they make it their specialty. In my view, the diagnosis and treatment of sleep disorders requires special training. Because of space constraints, I am not going to discuss the sleep disorders, and interested readers should consult the *DSM-5*.

Sexual Dysfunctions

Sexual dysfunctions include delayed ejaculation, erectile disorder, female orgasmic disorder, female sexual interest/arousal disorder, genito-pelvic pain/penetration disorder, male hypoactive sexual desire disorder, premature (early) ejaculation, and substance/medication-induced sexual dysfunction. This is another area that counselors typically do not diagnose unless they have specialized training in both the physiological and psychological aspects of sexual dysfunction. Once again, I am going to direct interested readers to the *DSM-5*.

L15 Gender Dysphoria

In this chapter there is one diagnosis, *gender dysphoria,* with separate criteria for children and adolescents and adults. There is considerable controversy on the terms *sex* versus *gender,* and in the *DSM-5* sex is related to biological factors, whereas gender denotes one's public lived role as a boy or girl, man or woman that is a combination of the interaction among biological, social, and psychological factors. For both children and adolescents/adults, gender dysphoria concerns "a marked incongruence between one's experienced/expressed

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

gender and assigned gender" (APA, 2013, p. 452). Clients must experience this incongruence for at least six months. There are more criteria that children must exhibit than adults, and children must meet the criterion of "a strong desire to be of the other gender or an insistence that one is the other gender (or some alternative gender different from one's assigned gender)" (APA, 2013, p. 452).

L16 Disruptive, Impulse-Control, and Conduct Disorders

This category of disorders involves problems with self-control of emotions and behaviors that violate the rights of others. The essential feature of *oppositional defiant disorder* is a pattern of angry/irritable mood, argumentative/irritable behavior, and vindictiveness. The symptoms of oppositional defiant disorder, such as "often loses temper," "is angry and resentful," and "often deliberately annoys others," must be present for at least six months. The next disorder in this chapter is *intermittent explosive disorder,* which is characterized as a failure to control aggressive impulses that results in behavioral outbursts. These aggressive outbursts are not premediated and are grossly out of proportion to the situation. Furthermore, they cause marked distress, impairment in occupational or interpersonal functioning, or financial or legal problems. As indicated earlier, the essential characteristic of *conduct disorder* is a repetitive and persistent pattern of behavior that violates the basic rights of others or violates major age-appropriate societal norms or rules.

Pyromania concerns the deliberate setting of fires where there is tension or arousal before the setting of the fire and pleasure, gratification, or relief either while setting the fire or experiencing the aftermath. *Kleptomania* involves the "recurrent failure to resist the impulses to steal objects that are not needed for personal use or for the monetary value" (APA, 2013, p. 478). Clients feel tension before stealing the object and then feel pleasure, gratification, or relief after the theft.

L17 Substance-Related and Addictive Disorders

Within this classification are three categories of diagnoses: *substance use disorder, substance withdrawal,* and *substance intoxication.* Also when clinicians diagnose within the substance-related disorder, there are separate code numbers dependent upon the substance that is problematic. The substances are (a) alcohol; (b) caffeine; (c) cannabis; (d) opioid; (e) hallucinogen (PCP); (f) inhalant; (g) sedative, hypnotic, or anxiolytic; (h) stimulant; (i) tobacco; and (j) other (or unknown). Thus, when counselors are diagnosing in the area they would code, for example, alcohol use disorder, alcohol withdrawal, alcohol intoxication. Table 15.3 lists the substance categories and the possible diagnoses.

As compared to withdrawal and intoxication, the most common diagnosis counselors will use is *substance use disorder* (e.g., opioid use disorder). The substance use disorders encompass two diagnoses from the *DSM-IV-TR* (i.e., substance dependence and substance abuse). To merit a diagnosis of substance use disorder, the client must have at least two of the symptoms listed here within the last 12 months.

1. Taking the substance in larger amounts or over a longer period than was originally intended.
2. A desire to cut down use and the client may report multiple times to cut down or stop taking the substance.
3. Spends a great deal of time obtaining, using, or recovering from the effects of the substance.
4. Craving or strong urge to use the substance.
5. Failure to fulfill major role obligations at work, school, or home.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 15.3
Substances and possible diagnoses

	Substance Use	Intoxication	Withdrawal
Alcohol	X	X	X
Caffeine		X	X
Cannabis	X	X	X
Hallucinogens			
Phencyclidine	X	X	
Other hallucinogens	X	X	
Inhalants	X	X	
Opioids	X	X	X
Sedatives, hypnotics, or anxiolytics	X	X	X
Stimulants	X	X	X
Tobacco	X		X
Other (or unknown)	X	X	X

6. Continued substance use despite having persistent or recurrent social or interpersonal problems.
7. Important activities are given up or reduced because of the substance use.
8. Recurrent substance use in situations where it is physically hazardous.
9. Continued substance use despite knowledge that there are physical or psychological problems related to substance use.
10. Tolerance.
11. Withdrawal (not a criterion for all substance).

As Table 15.3 reflects, substance use disorders are found in every classification of substances except for caffeine. Substance use disorders are measured on a continuum from mild to severe depending on the number of symptoms present.

Substance intoxication and withdrawal fall under the broader umbrella of substance-induced disorders as both of these diagnostic categories involve substance use. As Morrison (2014) indicated, intoxication involves the over use of the specific substance, and criteria for intoxication are included in each substance-specific section of the *DSM-5*. Criteria for substance withdrawal are also included in the specific substance sections of the *DSM-5*. The essential feature of withdrawal is the behavioral change, with physiological and cognitive concomitants, that is a result of either the cessation or reduction in heavy and sustained substance use.

In substance-related and addictive disorders, there is one non-substance-related disorder and that is *gambling disorder.* In gambling disorder, the client has persistent and recurrent problems with gambling behavior that results in clinically significant issues such as loss of relationships, friends, or job.

Neurocognitive Disorders

The neurocognitive disorders (known as "Delirium, Dementia, and Amnestic and Other Cognitive Disorders" in the *DSM-IV-TR*) are also an area of diagnosis that is rarely conducted by counselors. The neurocognitive disorders (NCDs) start with *delirium* and is followed by *mild neurocognitive disorders* and *major neurocognitive disorders,* which are followed by their etiological subtypes (e.g., major neurocognitive disorder due to Alzheimer's disease or mild frontotemporal neurocognitive disorder). A delirium is described as a disturbance

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

of attention and awareness and a change in cognition that develops over a short period of time. Delirium is usually caused by a medical condition or substance use and withdrawal. The neurocognitive disorders are unique in the *DSM-5* in that these disorders focus on cognitive deficits that have an underlying pathology that frequently can be identified (e.g., mild vascular neurocognitive disorder). Although many disorders have cognitive deficits (e.g., schizophrenia), only disorders whose core features are cognitive are included in this category. Furthermore, the cognitive impairments are acquired and represent a change in cognition; therefore, they are different from the neurodevelopmental disorders.

L18 Personality Disorders

Personality disorders are *enduring* patterns of inner experience and behaviors that are inflexible and maladaptive and that cause significant functional impairment or distress. An entirely new approach to diagnosing personality disorder was proposed by the *DSM-5* working group, but right at the end of the process, it was decided to continue with the method for diagnosing personality disorder used in the *DSM-IV-TR.* For readers interested in the assessment of personality disorder, this new method is described in Section III of the *DSM-5.*

Personality disorders are, in many ways, difficult to diagnose because the criteria are somewhat less precise and because the client, as a function of the personality disorder, is less able to report symptoms accurately (Fong, 1995). A personality disorder diagnosis can be present with or without another diagnosis. With a personality disorder, the client often perceives that others are the ones with the difficulties because their behavior and emotions are natural reactions to the situation. Because personality disorders are considered more difficult to diagnose, it is important for counselors to understand the general characteristics and early signs of these disorders. Clients with personality disorders will have long-term functioning difficulties. The onset of the pattern of inflexibility and maladaption is typically first observed in adolescence or early adulthood. Although a client may seek counseling for difficulties in reaction to a specific stressor or event, a counselor should determine if there is a chronicity of maladaption. Fong (1995) suggested that during the initial interview, counselors should ask clients questions about the duration of the problems and whether the client can recall other such periods of distress or difficulties. These enduring patterns of inner experience and behavior are markedly different from the expectations of the culture and are manifested in two or more areas: cognition, affectivity, interpersonal functioning, or impulse control. A personality disorder is not something that appears over a few months, nor is it a reaction to something catastrophic (e.g., living in a war-torn country); rather, it is an enduring part of the personality.

Clients' perceptions of their personality disorders are *egosyntonic,* meaning that the disorder is an integral part of the self. Clients with a personality disorder, on the other hand, usually perceive the disorder as not part of themselves. In other words, clients with depressive disorder will not see the depressive disorder as a part of who they are, but rather symptoms of their problems. Clients with a personality disorder, however, will see the maladaption as a part of themselves and view the situation as just being the way it is. Therefore, clients with personality disorders think counseling has a low probability of changing their situations because the problems are immutable.

Another important feature of personality disorders is significant impairment in social and/or occupational functioning. Although clients occasionally experience anxiety or depression, the most dominant features of personality disorders are impairments in occupational and social functioning (Fong, 1995). Because personality disorders are related to enduring traits, difficulties will arise across many different situations. One way the tendency toward inflexibility is manifested is that clients with personality disorders continue to use the same maladaptive strategies over and over again.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 15.4
DSM-5 clusters of personality disorders and diagnoses

Cluster DSM-5	Diagnosis
Cluster A	Paranoid Personality Disorder Schizoid Personality Disorder Schizotypal Personality Disorder
Cluster B	Antisocial Personality Disorder Borderline Personality Disorder Histrionic Personality Disorder Narcissistic Personality Disorder
Cluster C	Avoidant Personality Disorder Dependent Personality Disorder Obsessive-Compulsive Personality Disorder

A personality disorder may not be evident to a practitioner in the first counseling session, in which the focus is typically on the immediate or acute client difficulties. Fong (1995) proposed looking for the following signs that may be indicative of a possible personality disorder: (1) the counseling seems to suddenly stall or stop after making initial progress, (2) the client does not seem to be aware of the effect of his or her behaviors on others, (3) the client seems to accept the problems, (4) the client is underresponsive or noncompliant with the counseling regimen, and (5) the client is often involved with intense conflictual relationships with institutional or employment systems.

In the *DSM-5*, there are 10 personality disorders that are organized within 3 clusters (see Table 15.4). The clusters are atheoretical and are grouped together because of shared features. *Cluster A* concerns the disorders with odd and eccentric dimensions, whereas *Cluster B* includes disorders characterized by dramatic-emotional features. *Cluster C* contains the disorders with anxious-fearful characteristics.

Cluster A. Cluster A disorders are characterized by a client's lack of relationships, aloof behaviors, restricted affect, and peculiar ideas. There are three disorders within this cluster: paranoid personality disorder, schizoid personality disorder, and schizotypal personality disorder. The essential feature of *paranoid personality disorder* is a pattern of pervasive distrust and suspicion of others, such that others' behavior is considered threatening or malevolent. Clients with this disorder will expect other people (including the counselor) to exploit, harm, or deceive them. They do not trust others and have great difficulty with interpersonal relationships. These clients feel they have been deeply and irreversibly injured and are preoccupied with doubts about the loyalty or trustworthiness of friends and associates. They are continually looking for confirmation of their paranoid beliefs. Because they are constantly wary of harmful intentions, these individuals are often quick to identify any slight as an attack on their character or reputation. Furthermore, they often react with intense anger to any perceived insults.

The central feature of *schizoid personality disorder* is a pervasive pattern of detachment from social relationships and a restricted range of emotions. These individuals prefer to be alone and have little desire for personal relationships. They appear indifferent to the approval or criticism of others and are not particularly concerned about what other people may think of them. These individuals do not have close friends or confidants, except, possibly, a first-degree relative. They tend to have flattened affect or appear to be emotionally cold and detached. They are not likely to seek counseling because they see little purpose in talking with someone else or seeking assistance from another person.

The essential feature of the *schizotypal personality disorder* is a pervasive pattern of peculiar ideation and behavior with deficits in social and interpersonal relationships.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

These individuals incorrectly interpret casual incidents as having particular and unusual meanings to the individual. They have odd beliefs or magical thinking that is inconsistent with the subcultural norms (e.g., superstitions, belief in clairvoyance). Paranoia, unusual perceptions, and odd beliefs are evident, but they do not reach the level of chronic delusional proportions. They also have difficulties forming relationships with other people because their ideas and interpretations of current events often seem strange.

Cluster B. Individuals with Cluster B disorders are very different from Cluster A clients because the former are quite emotional and may often try to impress the counselor. The behavior of individuals in this group tends to be erratic and unstable, with affect that is quite changeable and heightened. The first of the four disorders in this cluster is the *antisocial personality disorder.* The central characteristic of the antisocial personality disorder is a pervasive pattern of disregard for and violation of others' rights, which begins in childhood or adolescence and continues into adulthood. This diagnosis, however, cannot be given until the person is at least 18 years of age and has a history of some of the symptoms of conduct disorder before the age of 15. After the age of 15, the individual's disregard for others is a pattern reflected in the following: (1) repeated involvement in illegal behaviors; (2) deceitfulness, lying, or conning others; (3) being impulsive and not planning; (4) aggressiveness and repeated physical fights or assaults; (5) reckless disregard for the safety of self and others; (6) being consistently irresponsible; and (7) a lack of remorse.

Borderline personality disorder is characterized by a pervasive pattern of instability in interpersonal relationships, self-image, and mood. This instability is also accompanied by impulsivity. Borderline clients may display frantic efforts to avoid real or imagined abandonment. Often their relationships are unstable and intense, fluctuating between idealizing and devaluing the other person. There is usually a marked and persistent disturbance of identity. There may be self-damaging behaviors and recurrent suicidal gestures or threats. These clients often have intense feelings, and there is considerable emotional instability. Others may perceive these individuals as overreacting with brief, but intense, episodes of depression, irritability, or anxiety. These individuals also have a tendency to feel chronically empty and have anger control problems. Furthermore, the basic dysphoric mood is often disrupted by feelings of anger, panic, or despair but is rarely interrupted by feelings of well-being or satisfaction. According to the American Psychiatric Association (2013), clients with borderline personality disorder make up about 10% of the individuals seen in outpatient mental health clinics; hence, there is a high probability that mental health counselors will work with clients that have a borderline personality disorder.

The essential features of the *histrionic personality disorder* are an excessive and pervasive emotionality and attention-seeking behaviors. These individuals are dissatisfied unless they are the center of attention. Their interactions with others may be inappropriately sexual or provocative. Emotions change rapidly, and their behavior is often considered inappropriately exaggerated, sometimes to the point of being theatrical. Their speech is dramatic and impressionistic but also tends to lack detail. They are often quite suggestible and perceive relationships to be more intimate than the relationships actually are. They often have difficulty maintaining work because they become easily bored and coworkers have difficulty with their emotionality and attention-seeking behavior.

With *narcissistic personality disorder,* there is a pattern of grandiosity, a need for admiration, and a lack of empathy. These individuals have a grandiose sense of self-importance and are preoccupied with their fantasies of success, brilliance, beauty, or ideal love. Narcissistic clients expect special regard from others but often devalue others' achievements and abilities. They require excessive admiration and expect to be catered to. They often exploit people and lack any empathy toward others. They feel entitled and expect to

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

receive acknowledgment for a job well done even though they may have contributed very little to project. Narcissistic clients are often envious of other people and expect that others are envious of them in return. They typically appear arrogant and seem to have haughty attitudes and behaviors. They are unlikely to seek counseling because any problematic situation is not related to their behavior, which they feel is beyond reproach.

Cluster C. The disorders in this cluster are characterized by the client's anxiety and avoidant behaviors. Clients with these disorders rigidly respond to demands by passively enduring, changing self, or withdrawing. The three disorders are the avoidant personality disorder, the dependent personality disorder, and the obsessive-compulsive personality disorder. The essential feature of the *avoidant personality disorder* is a pervasive pattern of social inhibition, feelings of inadequacy, and a fear of negative evaluation. These individuals avoid work, school, or even promotion opportunities because of these fears. They are unlikely to enter into relationships without strong guarantees of unrelenting acceptance. Because they are preoccupied with being criticized or rejected, they have a markedly low threshold for detecting such behaviors. They often view themselves as being socially inept, personally unappealing, or inferior to others. They show restraint in both relationships and activities because of fear that they may be ridiculed or embarrassed.

Dependent personality disorder is characterized by a pervasive and excessive need to be taken care of. This need leads to submissive and clinging behaviors accompanied by fears of separation. These individuals have great difficulty making decisions. They want others to take the lead and are fearful of disagreeing with them. Individuals with dependent personality disorder are fearful of being alone and will go to excessive lengths to obtain nurturance and support from others. If a relationship ends, they frantically seek another relationship as a source of care and support. They can go to excessive lengths to obtain nurturance and support from others.

The central feature of *obsessive-compulsive personality disorder* is a preoccupation with orderliness, perfectionism, and interpersonal and mental control. The person's overly stringent standards continually interfere with his or her ability to complete tasks or projects. Individuals with this disorder strive to make every detail perfect and display excessive devotion to work and productivity. They rarely take time for leisure, and when they do, the focus is on performing the leisure activity perfectly. Harsh judgments of self and others are common. Some people will have difficulty discarding even unimportant objects and may be frugal in their spending in order to be prepared for a future disaster. These individuals tend to be rigid and stubborn and contend that there is only a single "right" way to perform. Also with the general areas of personality disorders, there is *personality change due to another medical condition.*

L19 Paraphilic Disorders

The term *paraphilia* "denotes any intense and persistent sexual interest other than sexual interest in genital stimulation or preparatory fondling with phenotypical normal, physically mature, consenting human partners" (APA, 2013, p. 685). For *voyeuristic disorder,* the client must be at least 18 years old and over a period of at least six months have recurrent and intense sexual arousal from observing someone naked, in the process of disrobing, or having sexual activity, as manifested by fantasy, urges, or behaviors. In addition, either the person has acted on these fantasies with a nonconsenting person or these fantasies or urges have caused clinically significant distress or impairment. With *exhibitionistic disorder,* the client has fantasies, urges, or behaviors related to exposing one's genital to an unsuspecting person. Consistent with many other paraphilic disorders the urges last for at least six

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

months and are either acted upon or cause significant distress or impairment. *Frotteuristic disorder* concerns a persistent and intense sexual arousal from touching or rubbing against a nonconsenting person. These fantasies, urges, or behaviors last for six months and are acted upon or cause significant distress or impairment.

Sexual masochism disorder involves fantasies, urges, or behaviors related to intense sexual arousal from being humiliated, beaten, bound, or otherwise made to suffer. These sexual fantasies, urges, or behaviors last at least six months and cause clinically significant distress or impairment. With *sexual sadism disorder,* individual has fantasies, urges, or behaviors related to the infliction of physiological or psychological suffering of another person. These urges also must last for at least six months.

Pedophilic disorder involves recurrent and intense sexual arousing fantasies, sexual urges, or behaviors involving sexual activities with a prepubescent child or children. In this context, prepubescent is generally considered up to 13 years old. The individual has either acted on these fantasies, or these urges have caused marked distress or interpersonal difficulties. The individual needs to be at least 16 years old, and the child needs to be 5 years younger. The *DSM-5* notes that the clinician should not include an individual in late adolescence who is involved in an ongoing sexual relationship with a 12- or 13-year-old.

Fetishistic disorder concerns recurrent and intense sexual arousal from either the use of a nonliving object (e.g., shoes) or a highly specific focus on nongenital body part(s). Once again, these fantasies, urges, or behaviors need to last at least six months and cause clinically significant distress or impairment. With *transvestic disorder,* the intense sexual arousal comes from cross-dressing for at least six months.

Other Mental Disorders

There are four diagnoses listed in this section, which are rarely diagnosed by counselors. The diagnoses are *other specified mental disorder due to another medical condition, unspecified mental disorder due to another medical condition, other specified mental disorder,* and *unspecified mental disorder.*

L20 Other Conditions That May Be the Focus of Clinical Attention

This section covers conditions that are important in counseling; however, these conditions are not considered mental disorders. As a clinician, you list the numbers for these conditions because they are important to treatment. You can also list other factors that are not included in the *DSM-5* if there are factors that are important to the client's treatment. In the ICD-9, these conditions usually start with a V and are often referred to as V-codes. In the ICD-10, the conditions usually start with Zs. As these are not mental disorders, they are typically not reimbursed by insurance if they are listed alone.

In this section of the *DSM-5*, the first category is relational problems, and it includes such issues as *parent-child relational problems* and *disruption of family by separation or divorce.* The second category of codes concerns neglect and abuse, and it includes child maltreatment and neglect and adult maltreatment and neglect. Examples of conditions here are *child physical abuse* and *spouse or partner violence, physical.* There is also a category of education and occupational problems. The fourth category of codes focuses on housing and economic problems. This section of the *DSM-5* also includes a category entitled other problems related to social environment, which includes *acculturation difficulty.* Another section concerns problems related to crime or interaction with the legal system, which includes being both a *victim of a crime* and *conviction of a crime*. There are other categories and codes included in this section of the *DSM-5*, but hopefully you have gotten a gist of V-codes or Z-codes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Implications for Counselors

Now that we have discussed the disorders and other conditions that may be the focus of clinical attention that are contained in Section II of the *DSM-5*, it is important to review a few factors and discuss the implications for counselors. The *DSM-5* is grounded in the medical model, with a focus on diagnosis and treatment of a mental disorder, whereas counseling is steeped more in empowerment as indicated by the following definition: "Counseling is a professional relationship that empowers diverse individuals, families, and groups to accomplish mental health, wellness, education, and career goals" (Kaplan, Tarvydas, & Gladding, 2014). I would argue that these two approaches are not mutually exclusive and that diagnosis often plays a pertinent role in the working lives of many counselors. In fact, Eriksen and Kress (2006) argued that mental health counselors are diagnosing clients on a daily basis. I propose that it is possible to examine both deficits and strengths within one's approach to diagnosis.

The earlier discussion of the *DSM-5* is designed to be a primer. Interested counselors may also want to read Daily, Gill, Karl, and Barrio Minton (2014), which is designed as a learning companion to the *DSM-5* for counselors. As I have mentioned before, another excellent resource related to diagnosis is Morrison (2014). Interested readers should also thoroughly explore Section III of the *DSM-5*, which contains assessment measures, cultural formulation interviews, the alternative *DSM-5* model for personality disorders, and conditions for further study.

L21 Instruments Designed to Assist in Diagnosis

This section of the chapter is related to the application of assessment results in determining a diagnosis. Many assessment tools and instruments can assist the clinician in making this determination. Certainly, talking with and interviewing the client is one of the major methods for gathering information to determine a diagnosis. There are structured or semistructured clinical interviews that are designed specifically for diagnostic assessment. One of the more popular semistructured interviews is the *Structured Clinical Interview for DSM-5* (SCID-5), which was recently updated to correspond to the *DSM-5*. There are three versions of the SCID-5: the SCID-5-CV (Clinician Version), the SCID-5-RV (Research Version), and the SCID-5-CT (Clinical Trials Version). The SCID-5-CV is designed to be used as part of intake procedures in clinical settings and to help ensure a comprehensive diagnostic evaluation. A more comprehensive version of the SCID-5 is the SCID-5-RV (Research Version), which contains more disorders than the Clinician Version and includes all of the relevant subtypes, severity, and course specifiers. A feature of the SCID-5-RV is that it can be customized to fit the requirements of a particular study. Finally, the SCID-5-CT (Clinical Trials) is a modified version of the SCID-5-RV that has been streamlined for use in clinical trials research.

Another semistructured interview has been developed by the World Health Organization (WHO)—the World Mental Health-Composite International Diagnostic Interview (WMH-CIDI). The entire WMH-CIDI takes an average of two hours to administer in most populations. In total, there are 41 sections of the WMH-CIDI that not only address disorders but also functioning and physical disorders, current treatments, risk factors, and sociodemographics. The developers of the WMH-CIDI also developed elaborate training materials to teach interviewers how to administer this structured interview. As mental health services are needed throughout the world, it is important for counselors to be acquainted with instruments such as the WMH-CIDI that were developed to be used with international populations.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Millon Clinical Multiaxial Inventory

There are also standardized instruments developed for the specific purpose of assisting in the diagnostic process. One instrument, the *Millon Clinical Multiaxial Inventory-III* (MCMI-III; Millon, Davis, & Millon, 1997; Millon, 2006), is an interesting combination of a measure based on theory and a measure designed to correspond with the *DSM-IV.* There is a rumor that there will be a MCMI-IV, which will be released sometime soon. The MCMI-III is based on Millon's theory of personality, in which he suggested that psychopathology could be described by using two dimensions. The first dimension pertains to the primary source from which individuals gain positive reinforcement or avoid negative reinforcement, and the types are Detached, Discordant, Dependent, Independent, and Ambivalent. The second dimension concerns the client's coping behaviors, which can be either active or passive. These two dimensions produce 10 personality disorders (5 × 2), which are listed under Clinical Personality Patterns in Figure 15.2. The Severe Personality Patterns are a combination of types within the theory. These two categories also are consistent with the personality disorder diagnoses. The Clinical Syndromes and the Severe Clinical Syndromes, on the other hand, include many of the other disorders. The MCMI-III also includes indices of the examinee's attitudes in the Modifying Indices, which are similar to the Validity Scales of the MMPI-2.

The primary use of the MCMI-III is to differentiate psychiatric disorders, and this focus on differentiation is reflected in the scoring process. The scoring system uses what are called *base rates,* which calculate the probability that the client is or is not a member of a particular diagnostic group. The scoring of the MCMI-III is quite complex and can be done only by computer. Rather than using indicators such as two standard deviations above the mean to indicate psychopathology, the MCMI-III considers the actual prevalence of disorders in the clinical population. Prevalence rates do vary among settings, and some adjustments can be made in the scoring, depending on the setting. It is important to note, however, that an individual's responses on the MCMI-III are compared with the responses of psychiatric clients.

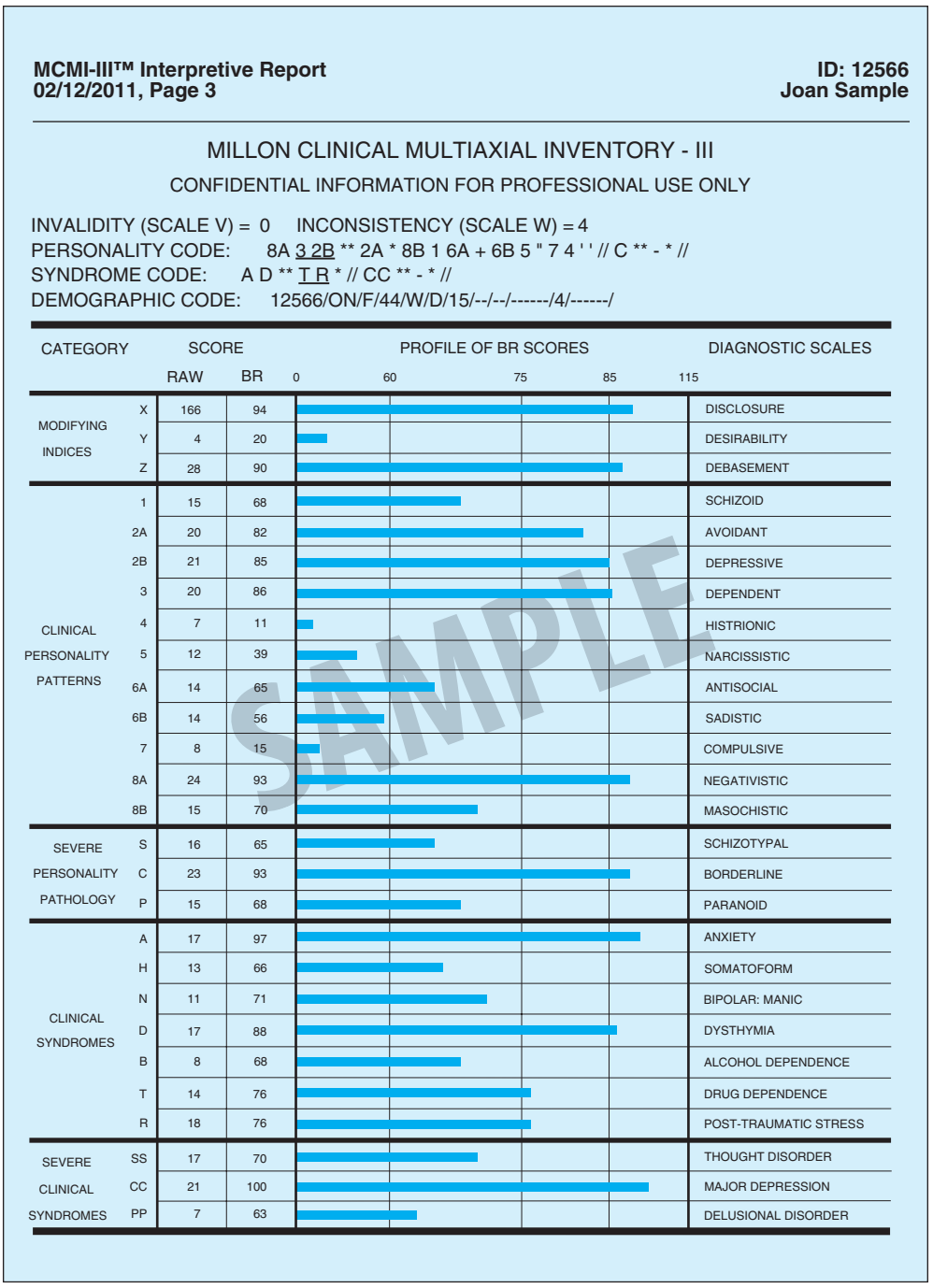
New to the interpretation of the MCMI-III are the Grossman Facet Scales. Grossman factor-analyzed the items of the MCMI-III and identified subcategories within each of the Clinical Personality Patterns and Severe Personality Pathology scales of the MCMI-III. Using these facet scales, trained clinicians can better pinpoint processes within the scales. For example, the Grossman Facet Scales for the Narcissistic scale are Admirable Self-Image, Cognitively Expansive, and Interpersonally Exploitive.

Counselors, therefore, should consider some important factors when using the MCMI-III. This instrument is *not* designed for normal functioning adults and should not be used as a general personality inventory. The MCMI-III should be used for diagnostic screening or clinical assessment. In addition, the computer-generated reports are written toward clients with moderate levels of pathology. Hence, for clients with lower ranges of psychological disturbance, the reports will probably overestimate their degree of difficulty. On the other hand, Fong (1995) recommended the use of the MCMI-III over other instruments with clients who may have symptoms of a personality disorder. There is empirical support for using the MCMI-III in diagnosing personality disorders (Craig, 1993). Millon's (1969) theory of personality functioning and psychopathology influenced the original formulation of Axis II personality disorders in earlier versions of the *DSM*. Most master's-level practitioners cannot use the MCMI-III unless supervised by someone with the appropriate qualifications.

Two other instruments have emerged from Millon's work. The first is the Millon Adolescent Clinical Inventory (MACI; Millon, Millon, & Davis, 1994), which also is

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 15.2
Profile of the Millon Clinical Multiaxial Inventory. From the Millon Clinical Multiaxial Inventory-III (MCMI-III).



MCMI-III™ Interpretive Report
02/12/2011, Page 3

ID: 12566
Joan Sample

MILLON CLINICAL MULTIAXIAL INVENTORY - III

CONFIDENTIAL INFORMATION FOR PROFESSIONAL USE ONLY

INVALIDITY (SCALE V) = 0 INCONSISTENCY (SCALE W) = 4
PERSONALITY CODE: 8A 3 2B ** 2A * 8B 1 6A + 6B 5 " 7 4 ' ' // C ** - * //
SYNDROME CODE: A D ** T R * // CC ** - * //
DEMOGRAPHIC CODE: 12566/ON/F/44/W/D/15/--/--/------/4/------/

CATEGORY		SCORE RAW	SCORE BR	PROFILE OF BR SCORES 0 60 75 85 115	DIAGNOSTIC SCALES
MODIFYING INDICES	X	166	94		DISCLOSURE
	Y	4	20		DESIRABILITY
	Z	28	90		DEBASEMENT
CLINICAL PERSONALITY PATTERNS	1	15	68		SCHIZOID
	2A	20	82		AVOIDANT
	2B	21	85		DEPRESSIVE
	3	20	86		DEPENDENT
	4	7	11		HISTRIONIC
	5	12	39		NARCISSISTIC
	6A	14	65		ANTISOCIAL
	6B	14	56		SADISTIC
	7	8	15		COMPULSIVE
	8A	24	93		NEGATIVISTIC
	8B	15	70		MASOCHISTIC
SEVERE PERSONALITY PATHOLOGY	S	16	65		SCHIZOTYPAL
	C	23	93		BORDERLINE
	P	15	68		PARANOID
CLINICAL SYNDROMES	A	17	97		ANXIETY
	H	13	66		SOMATOFORM
	N	11	71		BIPOLAR: MANIC
	D	17	88		DYSTHYMIA
	B	8	68		ALCOHOL DEPENDENCE
	T	14	76		DRUG DEPENDENCE
	R	18	76		POST-TRAUMATIC STRESS
SEVERE CLINICAL SYNDROMES	SS	17	70		THOUGHT DISORDER
	CC	21	100		MAJOR DEPRESSION
	PP	7	63		DELUSIONAL DISORDER

SAMPLE

aligned with the *DSM-IV-TR* diagnostic system. This instrument is appropriate for adolescents ages 13 through 19. The second instrument, the Millon Index of Personality Styles (MIPS; Millon, 1994), is not for diagnosis or clinical assessment; instead, it is intended for use with normal adults. This instrument incorporates Millon's theory with Jungian theory in an assessment of personality styles.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Summary

When counselors assess clients, their primary purpose is to gather information to effectively counsel clients. In determining the issues and problems that exist, the counselor will often, either formally or informally, diagnose the client. Hence, assessment, diagnosis, and treatment are often closely connected. Sound assessment information is the precursor to sound diagnostic decision making. This chapter focused on the most widely used diagnostic system in mental health settings in the United States, the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition* (DSM-5). Many of the concepts and instruments discussed previously in this book lead the clinician to a diagnosis, which then leads to treatment. It should be remembered, however, that not all clients have a mental disorder and therefore do not merit a *DSM-5* diagnosis. It should also be remembered that the *DSM-5* is a "living document" and that it will probably evolve and change. Therefore, counselors should stay abreast of diagnostic issues and systems.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

CHAPTER 16

Monitoring and Evaluating Counseling

Learning Objectives

- LO1 Understand the research supporting treatment monitoring
- LO2 Establish a method for monitoring treatment
- LO3 Distinguish between formative and summative evaluations
- LO4 Distinguish between qualitative and quantitative designs
- LO5 Identify what constitutes a randomized clinical trials design
- LO6 Identify outcome assessments to use in evaluating mental health services
- LO7 Identify outcome assessments to use in evaluating career interventions
- LO8 Identify outcome assessments to use in evaluating school counseling programs

You are a counselor in a community mental health center and you have been running a psychoeducational group for adolescents with behavioral problems (e.g., conduct disorder, oppositional defiant disorder). You are seeing remarkable changes in these adolescents, but you have no data that document this improvement. Your supervisor wants to change and expand the group to include many more members, which you believe will have a detrimental effect on the effectiveness of the group. You know your supervisor could be convinced to keep the group as you have it currently organized if you had convincing data to show her. Therefore, you need to collect data very quickly that document the effectiveness of your group.

According to Lambert and Vermeersch (2013), thousands of psychological tests have been published and most were specifically designed for the purposes of discrimination, prediction, or evaluation. In the previous chapters, we have discussed many instruments related to discrimination and prediction. A discriminative measure distinguishes between or among people on the basis of some underlying dimension or construct. For example, the Beck Depression Inventory-II was designed to discriminate between levels of depression and the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) discriminates between those with psychopathology and those without psychopathology. In terms of prediction, measures are used to classify people or predict future behavior. For example, the SAT is

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

designed for prediction. However, we have not yet talked specifically about using assessments in the important area of evaluation, which is the central theme of this chapter. Typically in evaluating counseling interventions, we are discussing an instrument that will measure client change (i.e., do clients get better as a result of the counseling).

In this discussion of evaluation, we will talk about evaluating clients' progress and counseling services. In these days of managed care, being able to monitor progress and show that clients are indeed getting better is a critical skill for all counselors.

LO1 Treatment Monitoring

To understand the research related to monitoring treatment, a counselor needs to understand some of the history of counseling and psychotherapy research. By the mid-1970s, research had clearly demonstrated that most people who received psychological interventions benefited from those services. There were, however, approximately 5 to 10% of clients who got worse during the course of therapy (Lambert, Bergin, & Collins, 1977). Unfortunately, there was little research interest in identifying those who got worse until the late 1980s when managed care began playing a significant role in cost containment. Although not used in a systematic manner in the 1990s, outcome assessment began playing a critical role in clinical care, and insurance companies became interested in identifying those clients who would not benefit from psychotherapy. At the same time, there were advances in statistical techniques in which statisticians became more sophisticated at developing growth models that could chart client trajectories (Lambert, 2010).

A few researchers began to investigate whether clinicians could identify those clients who would deteriorate. An interesting study by Hannan et al. (2005) found that the therapist could identify only 1 in 40 clients who deteriorated, whereas the computer program identified 36 out of the 40. Interesting new research, however, indicates that when clinicians receive feedback about client progress during psychotherapy, the clients have better outcomes than clients whose clinicians do not receive feedback (Whipple et al., 2003). That is, when counselors are informed that a client is not making progress or deteriorating, the counselors then adapt the process, which frequently results in a better outcome. Lambert, Hansen, and Finch (2001) used the Outcome Questionnaire-45 to monitor clients' weekly progress. By providing this simple feedback to clinicians, their clients improved significantly as compared to the clients whose therapists did not receive weekly monitoring information. The efficacy of providing counselors feedback was further substantiated by Lambert and colleagues (2003), who summarized the results from three large-scale studies. They found when clients were not improving and the counselor was given feedback regarding the lack of progress, the client started to make progress once the counselor received feedback. On the other hand, if clients were not making progress and business continued as usual, with counselors receiving no feedback, the clients often stayed the same. Their results indicated that feedback to the therapist makes an appreciable difference when clients are responding poorly to the counseling.

Even though client feedback, particularly when clients are not making progress, seems to have a positive influence on outcome, not all therapists appear to be open to receiving information about client progress. Bickman et al. (2000) found that only 23% of helping professionals wanted to receive outcome information more frequently than every fourth session. Lambert (2007) asserted that even good clinicians have better outcomes when information is shared that their clients are "off-track." For a review of the compelling research that documents the benefits of using assessments to monitor treatment progress, see Lambert (2010).

LO2 There are a number of treatment monitoring systems. One assessment-based method of treatment monitoring is MyOutcomes, which utilizes the Partners for Change Outcome

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Management System (PCOMS; Miller, Duncan, Sorrel, & Brown, 2005). The PCOMS is composed of two scales: one that focuses on assessing outcome and the other that assesses session factors. Each of these scales has only four items, so the PCOMS is quickly completed. Interested readers can visit https:// www.myoutcomes.com/ for more information on MyOutcomes. Another prominent system is CelestHealth System, which has a system for mental health agencies and university counseling center and a second system for medical care facilities. Users can select from either a 20- or 43-item assessment and also purchase the Psychotherapy Readiness Scale and the Therapeutic Bond Scale. Another frequently used system is the OQ-Analyst, which uses the *Outcome Questionnaire-45* (OQ-45; Lambert et al., 2004).

The OQ-Analyst is an Internet-based system where clients can securely logon to and take the OQ-45 and counselors can get immediate results. The OQ-45 is sensitive to client change and was designed so that it could be taken repeatedly in order to monitor clients' progress. The OQ-45 is designed to measure three domains of client functioning: (a) symptoms of psychological disturbance (particularly depression and anxiety), (b) interpersonal problems, and (c) social role functioning. The coefficient alpha for the total score is .93 and the test-retest is .84 (Lambert et al., 2004).

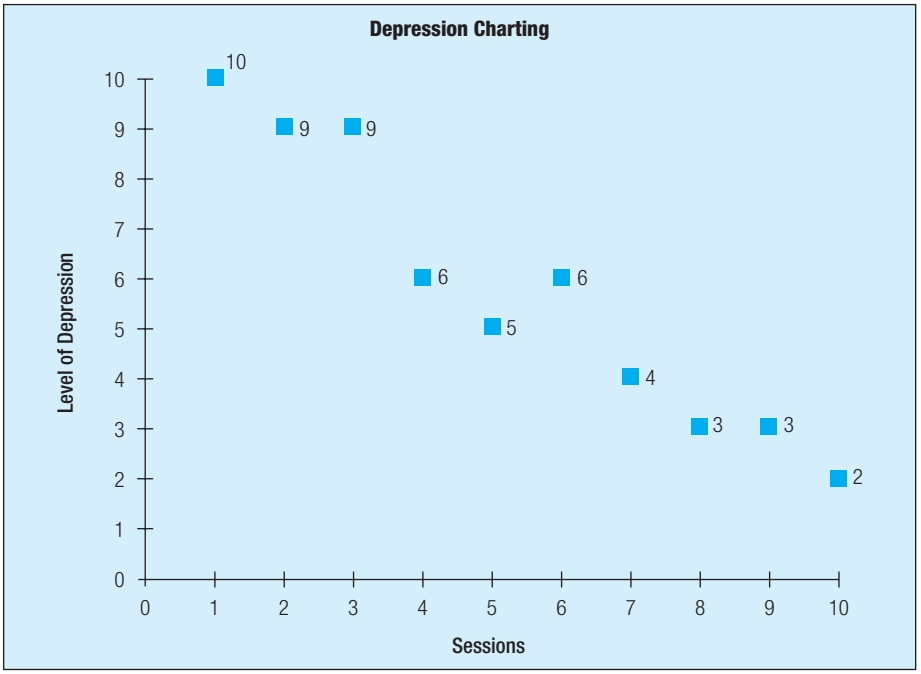
Another instrument that can be used in monitoring treatment is the Youth-Outcome Questionnaire (Y-OQ; Burlingame, Cox, Wells, Lambert, & Latkowski, 2005), which is for children 4 through 17. The Y-OQ is a 64-item instrument that is completed by the parent or guardian. Similar to the OQ-45, the Y-OQ was designed to track actual change in functioning as compared to assigning diagnoses. There are six subscales, which are Interpersonal Distress, Somatic, Intrapersonal Relations, Critical Items, Social Problems, and Behavioral Dysfunction. The Y-OQ Self-Report (Y-OQ-SR) is a similar measure except that it is a self-report measure completed by youth ages 12–17.

A key function in these treatment monitoring systems is detecting potential treatment failure. Research utilizing the OQ-45 indicates that the best predictors of outcome are initial severity of distress (i.e., pretreatment scores on the OQ-45) and change score in the first three sessions of treatment. Brown and Lambert (1998) found that pretreatment scores and change scores from Sessions 1 to 3 accounted for approximately 40% of the variance in final outcome. Therefore, knowing if change is occurring in the first few sessions of counseling is critical. Lambert and colleagues developed OQ-Analyst so that it could alert counselors to potential treatment failures. The OQ-Analyst program uses two predictive algorithms: (1) a statistical model based on a nationwide sample of over 11,000 cases and (2) a rational model based on clinician consensus ratings of satisfactory progress. Clinicians not only see a chart tracking the progress but also receive a message that varies depending on the degree of discrepancy between clients' results and the expected treatment response. The accuracy of this signal-alarm system has been evaluated in a number of studies and found to accurately predict 88% of cases with poor outcomes (Lutz et al., 2006; Percevic, Lambert, & Kordy, 2006; Spielmans, Masters, & Lambert, 2006).

When the OQ-Analyst system is activated, the clinician receives immediate feedback once the client has completed a session and taken the OQ-45. The feedback includes a graph of all OQ-45 total scores from pretreatment through the current session. Therefore, the counselors can see whether the client is improving. The counselor also receives a color-coded message (white, green, yellow, or red) that reflects the status of the treatment progress. White indicates the client is functioning in the normal range and that the counselor should consider termination. Green reflects that the client is making adequate progress and no change in treatment is recommended. Yellow suggests the client is making less than adequate progress and treatment suggestions are provided. Red indicates the client is not making the expected level of progress and encourages the counselor to carefully review this case and decide on a new course of action (e.g., referral for medication or intensification

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

FIGURE 16.1
Charting of client's self-report



of treatment). At every session, the counselor is able to review the graph and read the accompanying feedback. The OQ-Analyst system also includes the possibility of providing feedback directly to the client. Client feedback includes the graphing of total scores and messages that correspond to the therapists' feedback (i.e., white, green, yellow/red).

For counselors whose clients are not making adequate progress, they can also give the *Assessment for Signal Clients* (Lambert et al., 2008). The Assessment for Signal Clients is a 40-item measure that is designed to assess construct known to be related to outcome (e.g., quality of the therapeutic alliance, client motivation for change, client perceptions of social support). Specific feedback to the therapist is also provided based on these results. Research reflects that those therapists who receive the feedback and act on the clinical recommendations achieve significantly better outcomes with clients as compared to those clinicians who do not receive feedback (Lambert & Vermeersch, 2013).

Although there are many advantages to these treatment monitoring systems, counselors can also use more simplistic approach to treatment monitoring. For example, clinicians can systematically ask clients at the beginning of each session to rate their level of depression, anxiety, or other feelings on a scale of 1 to 10. Meir (2003) recommended that counselors graph these types of assessment to assist them in identifying trends and/or lack of progress. Figure 16.1 provides an example of charting a client's self-report on levels of depression on a scale of 1 to 10 over the course of 10 counseling sessions. In this case, the counselor could clearly see that progress was being made, and the client's level of depression was abating. Crucial to gathering a client's self-report of progress is a clear understanding by both the client and the counselor of what is being measured. For example, a counselor may have a client who has difficulties with insomnia analyze his or her level of anxiety. Although the counselor may see the direct link between the client's level of anxiety and insomnia, the client needs to understand the pertinence of the construct being assessed. Furthermore, with these types of self-assessments, there needs to be a clear definition of what the client is attempting to measure (e.g., what constitutes depression).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

If counselors want to make the continuous outcome assessment a little more formal yet still keep its individual or idiographic character, then they may want to consider *Goal Attainment Scaling* (GAS; Kiresuk, Smith, & Cardillo, 1994). The GAS has been used with a number of different client issues. The first step of this technique involves identifying a target area or complaint (e.g., level of depression). With GAS, the client and counselor need to select an indicator for each therapeutic goal. The indicator is the behavior, affective state, or process that most clearly represents the goal and can be used to indicate progress in reaching the goal. For example, with a goal of decreasing levels of depression, the indicator could be a depression scale or a record of the number of times in a week that the client feels "down." Avoid outcomes that are too vague or that could not be evaluated by an outside observer. Outcomes such as "better than when counseling began" are not specific enough. The next step is for the counselor and client to specify an expected level of outcome that is an accurate prediction of the client's status at the conclusion of treatment (e.g., feeling down only once or twice during a week). In Table16.1, these expected levels of outcome are associated with the numerical value of zero. Once the expected level of outcome has been determined, then the client and counselor determine levels for greater-than-expected outcomes and less-than-expected outcomes. In GAS, numerical ratings are associated with less-than-expected and greater-than-expected outcomes on a scale from −2 to +2 (see Table 16.1 for examples). There are also ways to combine scores across goals and convert them to standard scores if a clinician is doing research or wants group data (e.g., comparing effectiveness of different counselors).

Monitoring treatment progress can also involve some of the formal assessments discussed in Chapter 8 that measure client symptomatology. Some mental health organizations might consider using the Symptom Checklist-90—Revised (SCL-90-R; Derogatis, 1994) or the Brief Symptoms Inventory (BSI; Derogatis, 1993) as they are used during intake or in the initial session. These instruments have 90 and 53 items, respectively, and these assessments would be a quite cumbersome task for each session. An alternative might be the Brief Symptoms Inventory 18 (BSI-18), which has only 18 items and can be completed relatively quickly after every session.

TABLE 16.1
Example of Goal Attainment Scaling

Level of Attainment	Scale 1 Control of Anger	Scale 2 Career Planning	Scale 3 Insomnia
Much less than expected **–2**	Has an angry outburst once a day	Has done nothing	Within a week, has four nights with less than eight hours of sleep
Somewhat less than expected **–1**	Has an angry outburst two to six times a week	Has begun to look into other careers	
Expected level of outcome **0**	Has an angry outburst once a week	Has determined a new career and a plan for entering it	Within a week, has one night with less than eight hours of sleep
Somewhat less than expected **1**	Has an angry outburst two to five times a month	Has sent out résumés for a new job	
Much more than expected **2**	Has an angry outburst once a month	Has a new job	Sleeps eight hours every night for two weeks

COMMENTS

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

In conclusion, given the research regarding the importance of monitoring treatment, counselors are encouraged to institute a treatment monitor assessment. Whipple and Lambert (2011) purport that someday, treatment monitoring will be as common in psychotherapy as monitoring an individual's blood pressure is in medicine. Moreover, it may be important to share the results with the client, even when the results indicate that little progress is being made. This type of data can facilitate discussion of being "stuck" and consideration of what aspects of counseling are less productive for the client. Another advantage of treatment monitoring assessment is that it could also be used to gather accountability data that can be shared with different constituencies.

LO3 Evaluation in Counseling

As indicated previously, the second section of this chapter is devoted to evaluating counseling services. Counseling has entered a time when many stakeholders are interested in accountability information. For example, many insurance companies want data that documents that individuals do indeed get better from the services provided in a counseling agency. Furthermore, many counseling agencies also receive funding from nonprofit organizations (e.g., the United Way) or governmental agencies, and these organizations want data that shows that the dollars they invest will have tangible results. This is also true in school counseling, for the American School Counselor Association's (2012) National Model specifically encourages school counselors to gather accountability data that is measurable and linked to the school counseling programs. Counselors can never be quite sure when parents, administrators, oversight boards, or governmental agencies may want documentation that the services counselors provide are worth the monetary investment. Therefore, counselors need to understand the methods for gathering and analyzing accountability information.

Before I start discussing how to conduct an evaluation study, it is probably useful to discuss the distinctions typically made between the two major types of evaluation: formative evaluation and summative evaluation. **Formative evaluation** is the continuous or intermediate evaluation typically performed to examine the process. **Summative evaluation**, which is more cumulative than formative evaluation, concerns an endpoint or final evaluation. Sometimes the difference between the two is explained by the distinction that the focus of formative evaluation is on the *process*, whereas the focus of summative evaluation is on the *product*. The evaluation of counseling services for accountability reasons, which is more related to summative evaluation, is discussed later in this chapter.

Whiston (1996) advocated that practitioners provide accountability information by adapting the methodologies and measures developed in research of counseling outcomes. Outcome researchers conduct studies that indicate the degree to which counseling is effective. Often our interest in outcome research is determined by whether clients change (e.g., stop drinking or become less depressed) in a positive manner. A related concept is process research, which involves identifying factors or approaches that contribute to the outcome. Researchers have actively scrutinized the research methods for investigating the outcome of counseling for more than 60 years. This scrutiny has resulted in a wealth of information that counselors can use in evaluating their own effectiveness.

Defining Evaluation Study Focus

In developing evaluation or accountability information, clinicians first need to consider what they are attempting to evaluate. Sometimes, a counseling organization will seek one method of developing accountability information on all of the different services they provide. Identifying one method of developing accountability information may be difficult, however, because the services vary (e.g., individual counseling, group counseling, substance abuse services, parenting classes, brief crisis interventions). Thus, practitioners need

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

to determine what services they want to evaluate. As an example, an elementary school counselor may decide to focus first on evaluating a series of classroom guidance activities performed with fifth graders to ease the transition into middle school, rather than evaluating all the services the counselor provides to the children in that school. Because there needs to be a direct connection between the services provided and the outcome measures used, better accountability information can often be gathered by examining different types of counseling services separately.

L04 L05 Determining Design

Although there are a multitude of research designs, there are a few basic issues that a counselor needs to consider when developing an evaluation study. The first issue is whether the information sought from the evaluation study is more qualitative, quantitative, or both. **Qualitative data** is more descriptive, and in an evaluation study there are often open-ended questions or interviews where the clients can freely respond to the evaluation questions. With qualitative designs, the intent is to produce "rich" interpretative data. Usually with qualitative designs, the number of participants is comparatively small so that the counselor can get in-depth information from these participants. Conversely, **quantitative data** is more numerical in nature where the intent is to quantify the results. As an example, if the counselor wanted information on how to improve their counseling services, he might use a qualitative design and ask clients for suggestions. On the other hand, if the counselor was interested in knowing how many clients found the counseling services helpful, then a quantitative design would be best.

Frequently in evaluation studies in counseling, the question is whether the counseling services produce positive results for clients. Qualitative methodologies can be used in which clients describe their views of the changes that have occurred as a result of the counseling. Sometimes, however, counselors want quantitative data in which the goal is to produce some numerical indices of effectiveness. In quantitative studies of the effectiveness of a counseling program, the counselor needs to decide whether to use an intrasubject or intersubject design. In an *intrasubject* design, the same individuals are used for the comparison. A common approach is to give a pretest(s), provide the counseling interventions, and then assess the participants again with a posttest(s). The trouble with this approach is that participants' posttest scores can be influenced by other factors than solely the counseling interventions. Thus, many evaluators will select an *intersubject* design in which those who receive the counseling interventions are compared to an equivalent group except for the fact that they have not received the counseling intervention. A **randomized clinical trials design** is often considered the gold standard in intervention research. Adopted from the medical model, where patients are *randomly* assigned to receive the medication or the placebo, in counseling evaluation studies, clients are *randomly* assigned to either the intervention group or the placebo/control group. In most counseling situations, it is not ethical to simply assign clients who need services to a traditional placebo/control group where they do not receive any treatment. To address the ethical issue, what is frequently done is participants are randomly assigned to either the intervention group or the wait-list control group. Hence, both groups are given the pretest and then the intervention group receives the intervention and the wait-list control group waits until the intervention is completed and both groups then take the posttest. After the posttest has been administered, the wait-list control group then receives the intervention. Although randomized clinical trials are often considered the gold standard, they are not appropriate for all counseling situations (e.g., waiting for treatment for actively suicidal clients); hence, counselors should consider client needs when designing the evaluation study.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Selecting Participants

Once a counselor has determined the focus of the evaluation and the design of the study, the second step is to determine the clients who will participate. In some evaluation studies, counselors may want to use all clients to evaluate the program or services; at other times, a subset or sample of the client population may suffice. In terms of conducting an evaluation study, the old adage "more is better" often does apply to selecting a sample of participants. Of course for qualitative studies, the sample is usually smaller than for quantitative studies. For quantitative studies, a larger sample size does allow for more power in conducting statistical analyses, but many evaluation studies do not need to use sophisticated statistical analyses. Nevertheless, counselors should attempt to involve as many clients as is practical in the evaluation study in order to get various viewpoints and sound results.

Selecting Assessments or Measures

Many times, counselors will think that they need to develop a survey for an evaluation study; however, they should consider the concepts of reliability and validity and the need for a sound measure. As indicated earlier, in counseling evaluation studies, the process is quite similar to a counseling outcome research study and counselors can often use measures in which the reliability and validity has already been explored. The selection of outcome measures is, however, more intricate than simply selecting one measure that appears to be appropriate. Humans are complex and it is often difficult to assess the changes that may be occurring internally as a result of the counseling. Counseling is also complex and multidimensional; therefore, simple outcome measures may fail to adequately evaluate the therapeutic services provided. In addition, whose perspective should we use in determining outcome? Although we often ask the client if they are getting better, there may be times when a better indicator of actual change might come from a teacher or a parent. Interestingly, research has consistently shown that the effectiveness of counseling varies depending on whether the client, the counselor, or an outside observer completes the outcome measure (Orlinsky, Ronnestad, & Willutzki, 2004). Hence, measuring counseling outcome is not a simple enterprise; however, counselors can learn from the results of many years of substantial research on counseling outcomes. Two premier process and outcome researchers are Hill and Lambert (2004), who state that the best strategies in assessing outcome involve: (1) clearly specifying what is being measured; (2) measuring change from multiple perspectives (client, counselor, and outside observer); (3) using diverse types of assessments (e.g., rating scales, checklists); (4) using symptom-based and atheoretical measures; and (5) examining, as much as possible, patterns of change over time.

It is often difficult to select outcome measures, and counselors do not want to waste clients' time by having them complete two instruments that measure the same construct. Ogles, Lambert, and Fields (2002) developed a model for therapists and researchers to use to gather outcome data from a variety of viewpoints and domains (see Table16.2). The first dimension in their model is *content* and the subcategories are cognition, affect, and behavior. This would mean that if one outcome assessment measures feeling or affect, the clinician might want to consider another one that focuses on thinking or a behavioral rating scale. The second major category in the model is *social level*, which includes the categories of intrapersonal, interpersonal, and social role performance. In relation to this dimension, a practitioner may want to select instruments that have clients examine themselves (intrapersonal), their relationships with others (interpersonal), and how well they function in certain roles (social role).

The third dimension in the model is *source*, which includes self (i.e., the client), therapist, trained observer, and relevant other (e.g., parent or institution). An example of outcome that would fall under the institutional category is grade point average, in which the

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 16.2
Ogles, Lambert, and Fields' (2002) Scheme for Selecting Outcome Measures

Content	Social Level	Source	Technology	Time Orientation
Cognition	Interpersonal	Self	Global	Trait
a	*a*	*a*	*a*	*a*
b	*b*	*b*	*b*	*b*
Affect	Intrapersonal	Therapist	Specific	State
a	*a*	*a*	*a*	*a*
b	*b*	*b*	*b*	*b*
Behavior	Social Role	Trained Observer	Observation	
a	*a*	*a*	*a*	
b	*b*	*b*	*b*	
		Relevant Other	Status	
		a	*a*	
		b	*b*	
		Institutional		
		a		
		b		

outcome measure is retrieved from an institution (i.e., school). The fourth dimension is *technology*, or the type of assessment. The technology dimension includes the subcategories of global, specific, observation, or status. Some outcome assessments measure an overall level of functioning and would be categorized as global, whereas other measures may only assess level of anxiety and would be considered more specific. An observational assessment would be where the client is observed performing a desired outcome, such as a client with fear of flying being observed taking a flight. Status concerns a state or position and includes measures such as recidivism, expulsion from school, becoming divorced, or employment. The last dimension in Ogles et al.'s (2002) model is *time orientation*, which concerns whether the outcome measure is assessing a trait or a client state.

In using Ogles et al.'s (2002) model, a counselor does not need to find an outcome measure for each dimension or subcategory, as the goal is to consider comprehensive outcome assessment. Let us consider marital counseling and how a clinician may use the scheme to select outcome assessments. The counselor might consider the Marital Satisfaction Inventory—Revised, which in terms of content would be measuring affect. Hence, the clinician would consider other content areas such as behavior and possibly have the couple keep track of their arguments. The counselor might also consider having the couple complete a communication exercise at the beginning of counseling and at the end of counseling and observe different behaviors, which would include another source (i.e., therapist) and a different technology (i.e., observation). The Marital Satisfaction Inventory—Revised is a global measure, so the clinician may want to consider other types of technology (e.g., specific measures). An example of another technology measure could be status, such as documenting a couple's marital status at the end of the counseling (i.e., married, separated, or divorced).

As another example of considering multiple measures of outcome, let's consider a facilitator of a group for mildly depressed clients. To gain comprehensive information, the facilitator may begin by looking at the content subcategories of cognition, affect, and behavior. The clinician might consider using the Beck Depression Inventory-II (BDI-II), which would measure affect by having the clients chart behavioral aspects of depression

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

(e.g., sleep, eating, exercise). As the BDI-II measures intrapersonal affect, the clinician may want to consider some measure of interpersonal functioning. Social role might include developing a simple instrument to have clients evaluate their functioning at work. In terms of source, given that the group members are completing the BDI-II, the facilitator could complete the Hamilton Rating Scale for Depression (HRSD). The HRSD is designed for therapists to complete. In terms of technology, the BDI-II and the HRSD are specific to depression; the facilitator might want to consider a more global measure of outcome, such as the Outcome Questionnaire-45. In addition, the facilitator may want to consider a measure of satisfaction with the counseling.

In assessing outcome, multiple measures that draw relevant information from diverse sources often provide the best measures. Therefore, clinicians need to consider a variety of factors when selecting measures to use so they can gather sound accountability information. The following provides a brief description of some of the measures used in outcome research in mental health, career counseling, and school counseling settings.

LO6 ***Outcome Assessment in Mental Health Settings.*** As indicated earlier, providing mental health services to clients has been significantly influenced by the rise of managed care (Brown, Dreis, & Nace, 1999). Managed care agencies or third-party payers are interested in reimbursement for mental health services that are effective. They want to work with clinicians who can show that the counseling they provide results in clients getting better. Furthermore, many clients would respond positively if counselors could provide data showing that the majority of their clients do make significant improvements. In addition, clinicians need to continually analyze their own work to identify ways of providing the best possible service to clients. Hence, outcome data can provide critical information that serves clinicians in a variety of ways.

There has been interest in identifying a core battery of instrument to be used in outcome assessment; however, there are no established batteries, and different agencies use a variety of different instruments (Ogles, 2013). Thus, clinicians are often faced with the important duty of selecting outcome measures themselves. Commonly used instruments in research are the Beck Depression Inventory, author-created scales, diaries of behaviors or thoughts, State-Trait Anxiety Inventory, Symptom Checklist-90—Revised, Minnesota Multiphasic Personality Inventory-2, Dysfunctional Attitude Scale, Hassles Scale, and the Schedule for Affective Disorders and Schizophrenia (Hill & Lambert, 2004). Not all of these measures are good indicators of outcome as you want an instrument that is sensitive to change. For example, the MMPI-2 is not sensitive to change, and Lambert and Vermeersch (2013) recommend not using it as a measure of outcome. An instrument that is used in practice and research is the OQ-45, which was discussed earlier when we were talking about treatment monitoring. Another instrument that mental health clinicians might consider is the *Quality of Life Inventory* (Frisch, 1994), which measures satisfaction with 16 areas of life, such as work, love, and recreation. Another frequently used outcome measure is the *Satisfaction with Life Scale* (Diener, Emmons, Larsen, & Griffin, 1985). There also are a number of formal and informal measures of client satisfaction, such as the *Client Satisfaction Questionnaire 8* (Nguyen, Attkisson, & Stegner, 1983).

LO7 ***Outcome Assessments in Career Counseling.*** Similar to mental health counseling, there have been calls for a standard battery of instruments to be used in career counseling (Oliver & Spokane, 1988; Whiston, Sexton, & Lasoff, 1998); however, this has not yet occurred. Many of the research studies that have examined the efficacy of career counseling have involved measures of career maturity and career decidedness (Whiston, 2002), which were discussed in Chapter 11. Although these measures indicate some level of effectiveness,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

there are limitations, because many external agencies are more interested in concrete career outcomes (e.g., gaining employment) than in increases in career maturity or in becoming more decided in one's career choice. Whiston et al. (1998) also found variation among career outcome measures. They contended that measures of career decidedness typically showed little difference between those who received career counseling and those who did not. Hence, practitioners may want to consider other measures of effectiveness of career counseling than career maturity and career decidedness. Some noteworthy examples of career outcomes are employment, job satisfaction, and quality of life. In high school settings, counselors may want to consider academic measures and indicators of postsecondary planning. Lapan and Kosciulek (2001) presented a framework for identifying postsecondary placements for high school students and a procedure for creating career counseling programs with sensitivity to the needs of diverse communities.

With college students, there are issues related to how many times a student changes his or her major, time spent completing a degree, and grade point average, among others. Although Goal Attainment Scaling has rarely been used in career counseling research (Hoffman, Spokane, & Magoon, 1981), it has many advantages and is applicable in many career counseling situations.

There are also a few new measures that counselors should consider in the career area. Recently, Savickas and Porfeli (2011) revised the *Career Maturity Inventory*, which is a welcome addition to the family of career-related outcome measures. Savickas and Porfeli (2012) were also part of a large research team from 18 different countries that produced the *Career Adapt-Abilities Scale* (CAAS). The CAAS is composed of four scales that measure concern, control, curiosity, and confidence (see career construction theory; Savickas, 2005). The CAAS has been piloted in 13 countries: Belgium, Brazil, China, France, Iceland, Italy, Korea, the Netherlands, Portugal, South Africa, Switzerland, Taiwan, and the United States (Savickas & Porfeli, 2012). The CAAS demonstrated metric invariance across all the countries, and the reliability estimates were good to excellent for all of the countries. As there is increasing demand for instruments that can be used internationally, the CAAS is an excellent example of a theoretically based assessment that was designed from the beginning to be used in various countries.

LO8 ***Outcome Assessments in School Counseling.*** The American School Counselor Association's (ASCA, 2012) National Model, a document designed to guide school counseling programs, clearly states that school counseling programs are data driven. The ASCA National Model urges school counselors to move beyond simply gathering data on services that are provided to students to gathering data on the impact of these programs. According to the ASCA National Model, "To achieve the best results for students, school counselors regularly evaluate their program to determine its effectiveness. School counselors use this evaluation to answer the question, 'How are students different as a result of the school counseling program?' Now more than ever, school counselors are expected to demonstrate the effectiveness of their programs in measurable terms" (p. 99). In school counseling, as compared with other counseling areas, there has been significantly less outcome research (Whiston, Tai, Rahardja, & Eder, 2011). Furthermore, as noted by Whiston and Sexton (1998), the outcome instruments used in research are often author-developed for that particular study. Hence, the school counselor does not have the same supply of researched instruments to use in evaluating school counseling programs as mental health and career counselors have to evaluate their activities. Whiston et al. (2011) found that school counseling outcome researchers frequently used measures of self-esteem, which tend not to produce significant results. Whiston et al. suggested that school counselors not rely so frequently on measures of self-esteem and use other measures to evaluate school counseling activities and programs. School counselors do have access

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

TABLE 16.3
School Counseling Outcome Measures

Measures of Students' Performance/Attitudes
Grades or grade point average
Achievement test scores
Pursuing postsecondary training
Dropout rates
Credit completion rates
Behavioral ratings
Disciplinary referrals
Symptomatologies/adjustment measures
Career measures (e.g., career maturity)
Satisfaction with school counseling program/interventions
Measures of Teachers' Behavior/Attitudes
Satisfaction with school counseling program/interventions
Referral rate to school counselor
Ratings of student behaviors
Measures of Parents' Behavior/Attitudes
Satisfaction with school counseling program/interventions
Parenting measures
Ratings of student behaviors (e.g., homework time)
School involvement measures
Measures of Community Members' Behavior/Attitudes
Satisfaction with school counseling program/interventions
School involvement measures

to many important outcome assessments, such as measures of academic achievement (e.g., achievement test scores), discipline rates, and problem-solving skills.

In selecting outcome instruments, school counselors should also consider the general guidelines of using multiple measures from multiple perspectives. Table 16.3 contains a number of different strategies for evaluating school counseling interventions and programs, and it also shows that school counselors should consider gathering evaluative information from students, teachers, parents, and other members of the community.

As an example of the importance of continuing to gather accountability data, let's consider a school district that is experiencing some financial problems because the state has cut its budgetary allocations to education. This school district has four elementary schools, and each of these schools has a school counselor. However, because elementary school counselors are not mandated in this state, the school board is seriously considering eliminating these elementary school counseling programs. If these school counselors do not have compelling data indicating that their activities make a positive difference, then they can probably expect to lose their jobs. They would be more likely to retain their jobs if they have data indicating that their students have better achievement test scores, attend more school days, and have fewer discipline problems when compared with students in elementary schools that do not have elementary school counselors. Furthermore, if these school counselors had instituted an anti-bullying or a peer mediation program, then the school board might be more favorably disposed toward that program if reports existed showing that fights on the playground had decreased. Counselors often procrastinate in gathering accountability information because they are committed to and engaged in helping others. This commitment to others, however, needs to include gathering accountability information to ensure that needed counseling services for individuals and groups continue to be available to them.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Data Analysis

Some counseling practitioners shy away from conducting evaluation studies and gathering accountability information because they are unsure of how to analyze the data. New computerized statistical packages make data analysis less difficult and more user friendly than they have been in the past. Sometimes, descriptive information, rather than statistical analysis, is sufficient, with practitioners describing the changes in the outcome measures. Thompson and Snyder (1998) advocated that researchers should use effect size more often because of its practical applications. Effect size is traditionally calculated by subtracting the control group's mean from the experimental group's mean and dividing by the standard deviation of the experimental group (Glass, McGaw, & Smith, 1981), although there are other methods of calculating effect size (see Hedges & Olkin, 1985; Rosenthal, 1991). Sometimes, practitioners in the field do not have a control group. If this is the case, they can calculate average effect size by subtracting the pretest score from the posttest score and dividing by the posttest standard deviation. Practitioners should also be aware that many researchers are often willing to consult with practitioners on methodological and statistical questions. Therefore, if clinicians are unsure of the correct methods for analyzing the data, they should consult with someone who has expertise in that area.

Disseminate Results

Sometimes counselors conduct evaluation studies and may show the results to their immediate supervisor but they do not distribute the results widely. This is a mistake, particularly if the results are positive. Clients are more likely to invest in a process if they believe that the process will be helpful, and providing evaluation information to them can often facilitate the investment process. School counselors should be sharing results with the school board, parents, and the press. Counselors working in a community agency should also share their results with community stakeholders and the press. Sometimes an evaluation study simply stays in file drawers gathering dust when they could be informing people about the effectiveness of interventions.

In conclusion, there appears to be increased pressure for counselors to provide accountability information on their own effectiveness. The previous section discussed the steps in conducting an evaluation study, which are defining evaluation study focus, determining design, selecting participants, selecting assessments or measures, analyzing the data, and disseminating the results.

Summary

In this chapter, the discussion focused on moving beyond diagnosis to using assessment results in two primary ways: in monitoring client progress and for evaluation purposes. There is substantial research that supports the monitoring of clients' progress. Because early gains in counseling are particularly predictive of positive outcome, it is important to monitor changes from the beginning and identify clients who are not making progress. Monitoring client progress requires a delicate balance between gathering sound and comprehensive information and having procedures that are realistic and "doable." Furthermore, monitoring information must be readily available to the clinician and provided in a format that facilitates the identification of client progress or lack of improvement. Although monitoring client progress can be somewhat time consuming, the information provided is typically worth the time invested.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Another area in which assessment results play a crucial role in counseling is in evaluating the services provided to clients. Because of increasing demands for counselors to provide accountability information, many practitioners use assessment instruments to conduct evaluation studies. Evaluation study can be either formative or summative in design, which depends on the type of information the counselor is seeking. This chapter also included the steps for conducting an evaluation study and suggested a process for selecting measures, which was adapted from counseling outcome research. Moreover, some suggestions were made regarding evaluative assessments in the areas of mental health, career, and school counseling.

Concluding Remarks

In this book, I have argued that assessment is an integral part of the counseling process and that counselors need to be skilled assessors. To provide effective counseling services, counselors must first assess the client. For counselors, competency in both formal and informal assessment techniques is a necessity in today's mental health care environment. Often formal and informal assessment strategies can complement each other, resulting in assessments that are more effective. Informal assessments are subject to counselor bias and typically lack reliability and validation evidence. Formal assessment instruments, although typically objective, are often narrow in focus and cannot measure all client factors. The purpose of this book is to assist counseling students, counselors, and others in the helping professions to become skillful consumers of assessment measures. By using assessment measures effectively, practitioners can assist clients in making decisions, solving problems, functioning more effectively, and achieving their potential.

As I close this book, I want to encourage readers to consider their future clients. In terms of assessment skills, each counselor has strengths and limitations, and it is important for you to consider what your strengths and limitations are in this area. In considering your professional future, you need to look at the methods for expanding those strengths in assessment and identify avenues for remediating your limitations. All of your clients deserve professional and ethical counseling services, which can be facilitated by the development of competent assessment skills. Therefore, I want to encourage you to continue to expand your assessment skills.

Visit CengageBrain.com for a variety of study tools and useful resources such as video examples, case studies, interactive exercises, flashcards, and quizzes.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

APPENDIX A

2014 American Counseling Association Code of Ethics Section E

Section E: Evaluation, Assessment, and Interpretation

Introduction

Counselors use assessment as one component of the counseling process, taking into account clients' personal and cultural context. Counselors promote the well-being of individual clients or groups of clients by developing and using appropriate educational, mental health, psychological, and career assessments.

E.1. General

E.1.a. Assessments

The primary purpose of educational, mental health, psychological, and career assessment is to gather information regarding the client for a variety of purposes, including, but not limited to, client decision making, treatment planning, and forensic proceedings. Assessment may include both qualitative and quantitative methodologies.

E.1.b. Client Welfare

Counselors do not misuse assessment results and interpretations, and they take reasonable steps to prevent others from misusing the information provided. They respect the client's right to know the results, the interpretations made, and the bases for counselors' conclusions and recommendations.

E.2. Competence to Use and Interpret Assessment Instruments

E.2.a. Limits of Competence

Counselors use only those testing and assessment services for which they have been trained and are competent. Counselors using technology-assisted test interpretations are trained in the construct being measured and the specific instrument being used prior to using its technology-based application. Counselors take reasonable measures to ensure the proper use of assessment techniques by persons under their supervision.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

E.2.b. Appropriate Use

Counselors are responsible for the appropriate application, scoring, interpretation, and use of assessment instruments relevant to the needs of the client, whether they score and interpret such assessments themselves or use technology or other services.

E.2.c. Decisions Based on Results

Counselors responsible for decisions involving individuals or policies that are based on assessment results have a thorough understanding of psychometrics.

E.3. Informed Consent in Assessment

E.3.a. Explanation to Clients

Prior to assessment, counselors explain the nature and purposes of assessment and the specific use of results by potential recipients. The explanation will be given in terms and language that the client (or other legally authorized person on behalf of the client) can understand.

E.3.b. Recipients of Results

Counselors consider the client's and/or examinee's welfare, explicit understandings, and prior agreements in determining who receives the assessment results. Counselors include accurate and appropriate interpretations with any release of individual or group assessment results.

E.4. Release of Data to Qualified Personnel

Counselors release assessment data in which the client is identified only with the consent of the client or the client's legal representative. Such data are released only to persons recognized by counselors as qualified to interpret the data.

E.5. Diagnosis of Mental Disorders

E.5.a. Proper Diagnosis

Counselors take special care to provide proper diagnosis of mental disorders. Assessment techniques (including personal interviews) used to determine client care (e.g., locus of treatment, type of treatment, recommended follow-up) are carefully selected and appropriately used.

E.5.b. Cultural Sensitivity

Counselors recognize that culture affects the manner in which clients' problems are defined and experienced. Clients' socioeconomic and cultural experiences are considered when diagnosing mental disorders.

E.5.c. Historical and Social Prejudices in the Diagnosis of Pathology

Counselors recognize historical and social prejudices in the misdiagnosis and pathologizing of certain individuals and groups and strive to become aware of and address such biases in themselves or others.

E.5.d. Refraining From Diagnosis

Counselors may refrain from making and/or reporting a diagnosis if they believe that it would cause harm to the client or others. Counselors carefully consider both the positive and negative implications of a diagnosis.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

E.6. Instrument Selection

E.6.a. Appropriateness of Instruments

Counselors carefully consider the validity, reliability, psychometric limitations, and appropriateness of instruments when selecting assessments and, when possible, use multiple forms of assessment, data, and/or instruments in forming conclusions, diagnoses, or recommendations.

E.6.b. Referral Information

If a client is referred to a third party for assessment, the counselor provides specific referral questions and sufficient objective data about the client to ensure that appropriate assessment instruments are utilized.

E.7. Conditions of Assessment Administration

E.7.a. Administration Conditions

Counselors administer assessments under the same conditions that were established in their standardization. When assessments are not administered under standard conditions, as may be necessary to accommodate clients with disabilities, or when unusual behavior or irregularities occur during the administration, those conditions are noted in interpretation, and the results may be designated as invalid or of questionable validity.

E.7.b. Provision of Favorable Conditions

Counselors provide an appropriate environment for the administration of assessments (e.g., privacy, comfort, freedom from distraction).

E.7.c. Technological Administration

Counselors ensure that technologically administered assessments function properly and provide clients with accurate results.

E.7.d. Unsupervised Assessments

Unless the assessment instrument is designed, intended, and validated for self-administration and/or scoring, counselors do not permit unsupervised use.

E.8. Multicultural Issues/Diversity in Assessment

Counselors select and use with caution assessment techniques normed on populations other than that of the client. Counselors recognize the effects of age, color, culture, disability, ethnic group, gender, race, language preference, religion, spirituality, sexual orientation, and socioeconomic status on test administration and interpretation, and they place test results in proper perspective with other relevant factors.

E9. Scoring and Interpretation of Assessments

E.9.a. Reporting

When counselors report assessment results, they consider the client's personal and cultural background, the level of the client's understanding of the results, and the impact of the results on the client. In reporting assessment results, counselors indicate reservations that exist regarding validity or reliability due to circumstances of the assessment or inappropriateness of the norms for the person tested.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

E.9.b. Instruments With Insufficient Empirical Data

Counselors exercise caution when interpreting the results of instruments not having sufficient empirical data to support respondent results. The specific purposes for the use of such instruments are stated explicitly to the examinee. Counselors qualify any conclusions, diagnoses, or recommendations made that are based on assessments or instruments with questionable validity or reliability.

E.9.c. Assessment Services

Counselors who provide assessment, scoring, and interpretation services to support the assessment process confirm the validity of such interpretations. They accurately describe the purpose, norms, validity, reliability, and applications of the procedures and any special qualifications applicable to their use. At all times, counselors maintain their ethical responsibility to those being assessed.

E.10. Assessment Security

Counselors maintain the integrity and security of tests and assessments consistent with legal and contractual obligations. Counselors do not appropriate, reproduce, or modify published assessments or parts thereof without acknowledgment and permission from the publisher.

E.11. Obsolete Assessment and Outdated Results

Counselors do not use data or results from assessments that are obsolete or outdated for the current purpose (e.g., noncurrent versions of assessments/instruments). Counselors make every effort to prevent the misuse of obsolete measures and assessment data by others.

E.12. Assessment Construction

Counselors use established scientific procedures, relevant standards, and current professional knowledge for assessment design in the development, publication, and utilization of assessment techniques.

E.13. Forensic Evaluation: Evaluation for Legal Proceedings

E.13.a. Primary Obligations

When providing forensic evaluations, the primary obligation of counselors is to produce objective findings that can be substantiated based on information and techniques appropriate to the evaluation, which may include examination of the individual and/or review of records. Counselors form professional opinions based on their professional knowledge and expertise that can be supported by the data gathered in evaluations. Counselors define the limits of their reports or testimony, especially when an examination of the individual has not been conducted.

E.13.b. Consent for Evaluation

Individuals being evaluated are informed in writing that the relationship is for the purpose of an evaluation and is not therapeutic in nature, and entities or individuals who will receive the evaluation report are identified. Counselors who perform forensic evaluations obtain written consent from those being evaluated or from their legal representative unless a court orders evaluations to be conducted without the written consent of the individual being evaluated. When children or adults who lack the capacity to give voluntary consent are being evaluated, informed written consent is obtained from a parent or guardian.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

E.13.c. Client Evaluation Prohibited

Counselors do not evaluate current or former clients, clients' romantic partners, or clients' family members for forensic purposes. Counselors do not counsel individuals they are evaluating.

E.13.d. Avoid Potentially Harmful Relationships

Counselors who provide forensic evaluations avoid potentially harmful professional or personal relationships with family members, romantic partners, and close friends of individuals they are evaluating or have evaluated in the past.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

APPENDIX B

Table of Areas Under the Normal Curve

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0010	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0077	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	3336	.3300	.3264	.3228	.3194	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.0	5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

APPENDIX C

Standards for Multicultural Assessment*

AACE Standards for Multicultural Assessment (4th ed., 2012), Fourth Revision, 2012

The Association for Assessment in Counseling and Education (AACE) is an organization of counselors, educators, and other professionals that advances the counseling profession by providing leadership, training, and research in the creation, development, production, and use of assessment and diagnostic techniques. The mission and vision of AACE drives the continuing effort to create awareness of the importance of assessment and diagnostic techniques for an increasingly diverse population, to promote better training in the uses and development of assessments and to advocate for social justice concerns in counseling and educational assessment.

In 1992, the Committee on Diversity in Assessment under the direction of the Association for Assessment in Counseling (now AACE) Executive Council created the first set of standards addressing cultural diversity and the impact on assessment in counseling and education. In 2003, the Executive Council of the Association for Assessment in Counseling (AAC) appointed a committee to expand and update the original set of standards.

During the past decade, the Council for Accreditation of Counseling and Related Educational Programs (CACREP) emphasized the centrality of diversity in our profession by including language regarding diversity and advocacy throughout every aspect of the 2009 CACREP Standards. The Council on Rehabilitation Education (CORE) has also acknowledged in its 2010 Standards the importance of cultural and individual diversity and has incorporated cultural competence into standards for rehabilitation counseling programs.

This revision of the Multicultural Assessment Standards addresses more specifically the role of social advocacy in assessment. Further, they speak to the importance of effectively selecting, administering, and interpreting assessments and diagnostic techniques while providing training standards in multicultural assessment. The intent of the revision is to enhance counseling professionals' knowledge of as well as the public's awareness and support for culturally-appropriate assessment. This is particularly salient

*©Association for Assessment in Counseling. All rights reserved. Reprinted by permission.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

as assessment is an individual and system interventions useful for client/student and community empowerment, advocacy, collaboration, to change systems, and inform public opinion and policy.

Definition of Multicultural and Diverse Populations

We define multicultural populations broadly to recognize persons who differ by race, ethnicity, culture, language, age, gender, sexual orientation, religion, and ability. Today counselors consider during the assessment process the impact of age, color, culture, ability, ethnic group, gender, gender identity, race, religion, sexual orientation, linguistic background, socioeconomic status or other personal characteristics. Counselors are also challenged to advocate for culturally diverse individuals from marginalized populations. To this end, this revision of the Multicultural Assessment Standards offers an introductory section on Advocacy, however, we assert that advocacy be infused throughout the Standards.

The Standards

Advocacy

Culturally competent professional counselors recognize the importance of social justice advocacy; they integrate understanding of age, gender, ability, race, ethnic group, national origin, religion, sexual orientation, linguistic background, and other personal characteristics in order to provide appropriate assessment and diagnostic techniques.

Professional counselors should:

- Recognize in themselves and others, subtle biases and the way these biases influence and impact the assessment process for marginalized populations.
- Seek opportunities for learning by immersion into marginalized populations in order to gain understanding of clients' worldview and the impact on the assessment process.
- Support use of assessments with psychometric properties appropriate for individuals and vulnerable groups and create awareness about assessment of culturally diverse clients.
- Provide culturally competent and effective practices in all areas of counseling and assessment in individual, family, school, and community settings.
- Work collaboratively with community leaders to understand and address the needs of diverse clients providing opportunities to access services if needed.
- Address systemic barriers and consider how these barriers impact the interpretation and use of assessment results.
- Be knowledgeable of potential bias in assessment instruments and use procedures that comply with ethical guidelines when assessing marginalized populations.
- Are responsible for the appropriate applications, scoring, interpretations, and use of assessment instruments relevant to the needs of clients, whether they score and interpret such assessments themselves or use technology or other services.
- Take reasonable measures to ensure the proper use of psychological and assessment techniques by persons under their supervision, that results are kept confidential and that results are not misused by others.

Selection of Assessments: Content and Purpose, Norming, Reliability, and Validity

Culturally competent professional counselors select assessments and diagnostic techniques that are appropriate and effective for diverse client populations.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Professional counselors should:

- Understand and follow the standard development procedures utilizing accepted research methods when developing new instruments to ensure technical quality of the content domains evaluated.
- Be knowledgeable of the technical aspects of assessments, including the importance of reliability, validity, measurement error and scores and norms when selecting assessments.
- Evaluate representative samples of test questions or practice tests, directions, answer sheets, manuals, and score reports before selecting a test.
- Understand how to review information on the performance of test takers of diverse subgroups, using appropriate norming information to ensure adequate subgroup analyses.
- Understand how to select and utilize appropriate modified forms of tests for test takers with disabilities who need special accommodations.
- Select assessments that help identify client needs, strengths and resources for client empowerment and self-advocacy.
- Select instruments with which they are trained and are competent to use and adhere to the ethical standards for the administration, scoring, interpretation, or reporting procedures and ensure that persons under their supervision are aware of these standards.
- Recognize the impact of cultural identity on test administration and interpretation, and place test results in proper perspective with other relevant factors.

Administration and Scoring of Assessments

Culturally competent professional counselors recognize challenges inherent in assessment of persons and seek to provide administration and scoring of assessment to clients respecting age, gender, ability, race, ethnic group, national origin, religion, sexual orientation, linguistic background, and other personal characteristics.

Professional counselors should:

- Obtain, understand and follow all established protocol for administering standardized test instruments to clients with diverse linguistic backgrounds utilizing a competent and bilingual translator to address cultural identity throughout the assessment process including informed consent or other procedural directives.
- Become familiar with test question format and procedures for answering test questions as well as understand additional materials or needed equipment.
- Understand the nature and importance of test security, copyright, and ramifications of cheating and maintain confidentiality of scores using adequate security procedures.
- Be knowledgeable with procedures, materials and directions for scoring tests and/or monitoring scoring process to ensure accuracy of test scores, promptly reporting any errors and communicating corrected results promptly.
- Administer assessments only within the context of a defined professional relationship and utilize tests for their recommended purpose only.
- Obtain informed consent from clients regarding the nature and purpose of assessment methods to be used and ensure that the test taker has the opportunity to become familiar with test question format.

Interpretation and Application of Assessment Results

Culturally competent professional counselors acknowledge the importance of social justice advocacy in interpretation and communication of assessment results with diverse populations.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Professional counselors should:

- Understand the normative groups, technical information, benefits and limitations of assessments and potential for bias in scoring and interpretation of assessment instruments.
- Provide a holistic appraisal of clients by creating a strong foundational evaluation respecting the influence of culture, background and individual characteristics.
- Understand the influence of culture, background, and individual characteristics when designing and implementing interventions to achieve effective counseling outcomes.
- Recognize how the effects of stigma, oppression, and discrimination impact the interpretation and application of assessment results for culturally diverse clients.
- Recognize and collaborate with others to eliminate biases, prejudices, and discriminatory contexts in conducting evaluations, interpretations and providing interventions.
- Explain the nature and purpose of assessment and specific use of results in an understandable, developmental level of the client or the client's legally authorized representative providing information about the impact of culture on assessment results and interpretation.
- Consider other factors present in the client's situation (e.g., disability or cultural factors or systematic or internalized oppressing) before making any recommendations, when relevant.
- Do not use data or results from assessments that are obsolete or outdated and make every effort to prevent the misuse of obsolete measures and assessment data by others.
- Release assessment data in which clients are identified only with the consent of clients or their legal representatives, or court order and only released to professionals recognized as qualified to interpret the data.

Training In The Uses Of Assessments

Culturally competent professional counselors seek training and supervised experience to ensure they provide appropriate assessment and diagnostic techniques for diverse client populations.

Professional counselors should:

- Understand the tests characteristics and receive guidance on the levels of skills, knowledge and training required to administer.
- Understand individual and group approaches to multicultural assessment.
- Receive training in how assessment data can be used to develop client action plans to make systemic changes that benefit disadvantaged individuals and vulnerable groups.
- Gain knowledge in how to collaborate with allies and policy makers to develop assessment practices that empower clients and educate the general public about culturally appropriate assessment of culturally diverse individuals and groups.
- Engage in continuous education to increase knowledge of assessment and enhance job performance.
- Have training and expertise in the use of traditional assessment and testing instruments including technical aspects of the instruments as well as cultural limitations to use test instruments for the welfare of culturally different clients.

References

American Art Therapy Association (2011). *Art Therapy Multicultural/Diversity Competencies*. Retrieved from http://www.arttherapy.org/upload/multiculturalcompetencies2011.pdf

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

American Counseling Association (2005). *Code of Ethics and Standards of Practice of the American Counseling Association*. Retrieved from http://www.counseling.org/resources/codeofethics/TP/home/ct2.aspx

Association for Lesbian, Gay, Bi-Sexual, Transgender Issues in Counseling (2010). *Competencies for Counseling LGBTQ Clients and Competencies for Counseling Transgender Clients*. Retrieved from http://www.algbtic.org/resources/competencies

Association for Multicultural Counseling and Development (1996). *AMCD Multicultural Counseling Competencies*. Retrieved from http://www.multiculturalcounseling.org/

American Psychological Association (2004). *Code of Fair Testing Practices in Education*. Retrieved from http://www.apa.org/science/programs/testing/fair-testing.pdf

Council on Accreditation of Counseling and Related Educational Programs (2009). *2009 Standards*. Retrieved from http://www.cacrep.org/doc/2009%20Standards%20with%20cover.pdf

Counselors for Social Justice Position Statement on Academic Achievement Gap and Equity on Educational Services (2008). Retrieved from http://counselorsforsocialjustice.com/CSJ_Position_-_Academic_Achievement_Gap.pdf

Council on Rehabilitation Education (2012). *Accreditation Manual for Masters' Level Rehabilitation Counselor Education Programs*. Retrieved from http://www.corerehab.org/Files/Doc/PDF/COREStandardsPrograms.pdf

Commission on Rehabilitation Counselor Certification (2009). *Code of Professional Ethics for Rehabilitation Counselors*. Retrieved from http://www.crccertification.com/filebin/pdf/CRCCodeOfEthics.pdf

Lewis, Arnold, House, and Toporek (2003). *ACA Advocacy Competencies*. Retrieved from http://www.counseling.org/resources/competencies/advocacy_competencies.pdf

National Association of Alcoholism and Drug Abuse Counselors (2011). *Code of Ethics*. Retrieved from http://www.naadac.org/membership/code-of-ethics

National Association of School Psychologists (2010). *National Association of School Psychologists' Model for Comprehensive and Integrated School Psychological Services*. Retrieve from http://www.nasponline.org/standards/2010standards/2_PracticeModel.pdf

National Board for Certified Counselors (2005). *Code of Ethics*. Retrieved from http://nbcc.org/Assets/Ethics/nbcc-codeofethics.pdf

National Career Development Association (2010). *Career Counselor Assessment and Evaluation Competencies*. Retrieved from http://associationdatabase.com/aws/NCDA/asset_manager/get_file/18143/aacencda_assmt_eval_competencies

Association for Assessment and Counseling (2003). *Responsibilities of Users of Standardized Tests*. Retrieved from http://www.theaaceonline.com/rust.pdf

Texas Professional Educational Diagnosticians Board of Registry (2010). *Best Practice Guidelines*. Retrieved from http://regped.com/pdf/TPEDBestPract.pdf

2011–12 Executive Council

- President: Danica G. Hays, Old Dominion University, dhays@odu.edu
- President-Elect: Carl Sheperis, Lamar University, csheperis@gmail.com
- Past President: Joshua Watson, Mississippi State University—Meridian, jwatson@meridian.msstate.edu
- Secretary: Casey Barrio-Minton, University of North Texas, casey.barrio@unt.edu
- Treasurer: Stephanie Crockett, Oakland University
- Governing Council Representative: Joshua Watson, Mississippi State University—Meridian, jwatson@meridian.msstate.edu
- MAL Publications: Dale Pietrzak, University of South Dakota, dale.pietrzak@usd.edu

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

- MAL Awards: Susan Carmichael, susan.eaves@gmail.com
- MAL Membership: Amy McLeod, Argosy University—Atlanta, amymcleod1@gmail.com
- Graduate Student Representative: Jayne Smith, Old Dominion University, j5smith@odu.edu

Multicultural Assessment Standards Revision Task Force: Dr. Linda Foster, Chair (Foster_lh@mercer.edu, Mercer University); Members: Dr. Gabriel Lomas (lomasg@wcsu.edu), Danica Hays (DHays@odu.edu), Michael Becerra (mbecerra@bamaed.ua.edu), and Anita Neuer Colburn (aneuer@regent.edu)

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Glossary

Accessibility The notion that all examinees should have an unobstructed opportunity to demonstrate their capabilities on the construct(s) being assessed.

Accommodations An action taken in response to an individual with a disability in which there is a departure from the standard testing protocol in an attempt to adjust for the disability.

Acculturation A process of change and adaption that individuals undergo as a result of contact with another culture.

Achievement test An assessment in which the person has "achieved" knowledge, information, or skills through instruction, training, or experience. Achievement tests measure acquired knowledge and do not make any predictions about the future.

Adaptation The change in original instrument in terms of design or administration to increase the accessibility to individuals (e.g., visually impaired, limited English proficiency).

Affective instrument An instrument that assesses interest, attitudes, values, motives, temperaments, and the noncognitive aspects of personality.

Age or grade equivalent scores Scores used to compare individuals with other individuals at the same age that are calculated by item response theory or by using a norm-referenced approach.

Alternate or parallel forms Two forms of an instrument that can be correlated, resulting in an estimate of reliability.

Analogue observation In this type of observation, the counselor creates a simulated environment that is reflective of the client's natural environment.

Appraisal Another term for *assessment.*

Aptitude test A test that provides a prediction about the individual's future performance or ability to learn based on his or her performance on the test. Aptitude tests often predict either future academic or vocational/career performance.

Assessment A procedure for gathering client information that is used to facilitate clinical decisions, provide clients with information, or for evaluative purposes.

Authentic assessments Performance assessments that involve the performance of "real" or authentic applications rather than proxies or estimators of actual learning.

Behavioral assessment An assessment method in which the focus is typically on observing and recording of the precise behaviors of the client.

Bias testing A term that refers to the degree to which construct-irrelevant factors systematically affect a specific group's performance.

Cluster sampling A technique that involves using existing units or clusters rather than selecting individuals.

Coaching It involves longer training or practice on questions that are the same as or similar to the items on the test.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Coefficient of determination This statistic estimates the percent of shared variance between two sets of variables that have been correlated. The coefficient of determination (r^2) is calculated by squaring the correlation coefficient.

Common Core State Standards A set of academic standards that are designed to develop rigorous, consistent, and real-world learning goals that would ensure that all students, regardless of where they lived, would graduate from high school prepared for college and career.

Computer adaptive testing (CAT) A testing in which the computer adapts the next question for the student based on his or her response.

Concurrent validity A type of validation evidence in which there is no delay between the time the instrument is administered and the time the criterion information is gathered.

Conditional standard errors of measurement A type of standard error of measurement that takes into account the different score levels.

Construct irrelevance The degree to which scores or results are affected by factors that are extraneous to the instrument's intended purpose.

Construct underrepresentation The degree to which the instrument is unable to capture significant aspects of the construct.

Construct validity One of the three traditional forms of validity that is broader than either content or criterion-related validity. Many experts in assessment now argue that evidence of construct validity, which includes the other traditional forms of validity, applies in all types of psychological and educational assessment. This type of validation involves the gradual accumulation of evidence. Evidence of construct validity is concerned with the extent to which the instrument measures some psychological trait or construct and how the results can be interpreted.

Content-related validity One of the three traditional forms of validity in which the focus was on whether the instrument's content adequately represented the domain being assessed. Evidence of content-related validity is often reflected in the steps the authors used in developing the instrument.

Convergent evidence Validation evidence that indicates the measure is positively related with other measures of the construct.

Correlation coefficient A statistic that provides an indication of the degree to which two sets of scores are related. A *correlation coefficient* (r) can range from —1.00 to +1.00 and, thus, provides an indicator of both the strength and direction of the relationship. A correlation of +1.00 represents a perfect positive relationship; a correlation of —1.00 represents a perfect negative or inverse relationship. A correlation coefficient of .00 indicates the absence of a relationship.

Correlational method A statistical tool often used in providing validation evidence related to an instrument's relationship with other variables.

Criterion-referenced instrument Instruments designed to compare an individual's performance to a stated criterion or standard. Often criterion-referenced instruments provide information on specific knowledge or skills and on whether the individual has "mastered" that knowledge or skill. The focus is on what the person knows rather than how he or she compares to other people.

Criterion-related validity One of the three traditional types of validity in which the focus is the extent to which the instrument confirms (concurrent validity) or predicts to (predictive validity) a criterion measure.

Cronbach's alpha or coefficient alpha It is one of the methods of estimating reliability through the examination of the internal consistency of the instrument. This method is appropriate when the instrument is *not* dichotomously scored, such as an instrument that uses a Likert scale.

Decision theory A method that examines the relationship between an instrument and a criterion or predictor variable, which usually involves an expectancy table. Expectancy tables frequently are used to determine cutoff scores or to provide clients with information regarding the probability of a certain performance on the criterion that is based on scores on the assessment.

Differential item functioning (DIF) A set of statistical methods for investigating item bias that examines differences in performance among individuals who are equal in ability but are from different groups (e.g., different ethnic groups).

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Discriminant evidence Validation evidence that indicates the measure is not related to measures of different psychological constructs.

Domain sampling theory Another term for *generalizability theory*.

Duty to warn Most states have laws that either require or permit mental health practitioners to disclose information about client when that client is going to harm someone else.

Event recording One of the methods used in behavioral assessment, where the counselor records the number of times a target behavior or behaviors occur during a specified time period.

Expectancy table A method of providing validity evidence that involves charting performance on the criterion based on the instrument's score. It is often used to predict who would be expected to fall in a certain criterion category (e.g., who is likely to succeed in graduate school) and to determine cutoff scores.

Factor analysis A term that covers various statistical techniques that are used to study the patterns of relationship among variables with the goal of explaining the common underlying dimensions (factors). In assessment, factor analysis is often used to examine if the intended internal structure of an instrument is reflected mathematically. For example, a researcher would analyze whether all items on each subscale "load" with the other items on the appropriate subscale (factor) and not with another factor.

False negative In decision theory, a term used to describe when the assessment procedure is incorrect in predicting a negative outcome on the criterion.

False positive In decision theory, a term used to describe when the assessment procedure is incorrect in predicting a positive outcome on the criterion.

Formative evaluation A continuous or intermediate evaluation typically performed to examine the counseling services process.

Frequency distribution A chart that summarizes the scores on an instrument and the frequency or number of people receiving that score. Scores are often grouped into intervals to provide an easy-to-understand chart that summarizes overall performance.

Frequency polygon A graphic representation of the frequency of scores. The number or frequency of individuals receiving a score or falling within an interval of scores is plotted with points that are connected by straight lines.

General ability test Another term for *intelligence test*.

Generalizability theory An alternative model to the true score model of reliability. The focus of this theory is on estimating the extent to which specific sources of variation under defined conditions influence scores on an instrument.

Grade equivalent norms Norms that are typically used in achievement tests and provide scores in terms of grade equivalents. In some instruments, grade equivalent scores are not validated on each specific grade but are extrapolated scores based on group performance at each grade level.

High-stakes testing A type of testing where the outcome of such tests have significant consequences (e.g., high school graduation examination).

Histogram A graphic representation of the frequency of scores in which columns are utilized.

Hypomanic episode Characterized by similar symptoms as a manic episode except that it lasts for only *four days*.

Individualized Education Plan (IEP) An educational plan that is developed for each student who is receiving special education and related services. The plan is developed by a team of educators and the child's parents or guardians.

Instrument An assessment tool that typically is not related to grading. In this book, instruments include tests, scales, checklists, and inventories.

Intelligence tests Instruments that are designed to measure the mental capabilities of an individual. These assessments are also referred to as *general ability tests*.

Intercepts In regression, it is the constant in the regression equation or it is where the line crosses the y-axis.

Interrater reliability A measure used to examine how consistently different raters evaluate the answers to the items on the instrument.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Interval recording An assessment method that focuses on whether specific behavior(s) occur within a certain interval of time. It is also referred to as *time sampling, interval sampling,* or *interval time sampling.*

Interval scale A type of measurement scale in which the units are in equal intervals. Many of the statistics used to evaluate an instrument's psychometric qualities require an interval scale.

Item analysis An analysis that focuses on examining and evaluating each item within an assessment.

Item difficulty An item analysis method in which the difficulty of individual items is determined. The most common item difficulty index (p) is the percentage of people who get the item correct.

Item discrimination A form of item analysis that examines the degree to which an individual item discriminates on some criterion. For example, in achievement testing, item discrimination would indicate whether the item discriminates between people who know the information and people who do not.

Item response theory (IRT) A measurement approach in which the focus is on each item and on establishing items that measure the individual's ability or level of a latent trait. This approach involves examining the item characteristic function and the calibration of each individual item.

Kuder-Richardson formulas Two formulas (KR 20 and KR 21) that were developed to estimate reliability. Both of these methods are measures of internal consistency. KR 20 has been shown to approximate the average of all possible split-half coefficients. KR 21 is easier to compute, but the items on the instrument must be homogeneous.

Latent trait theory Another term for *item response theory.*

Major depressive episode A period in which the client is in a depressed mood or has lost interest or pleasure in nearly all activities for at least two weeks.

Manic episode A distinct period of abnormally elevated, expansive, or irritable moods and abnormally increased goal-directed activity or energy that lasts at least *one week* and is present most of the day. It is characterized by a feeling of becoming so driven that it causes marked impairments in occupational functioning, social activities, or relationships.

Mean The arithmetic average of the scores. It is calculated by adding the scores together and dividing by the number in the group.

Median The middle score, with 50% of the scores falling below it and 50% of the scores falling above it.

Mental disorder A syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning.

Mental Measurements Yearbooks (MMY) A series of yearbooks that contain *critiques* of many of the commercially available psychological, educational, and career instruments.

Mental status examination An examination used to describe a client's level of functioning and self-presentation. It is generally conducted during the initial session or intake interview and is a statement of how a person appears, functions, and behaves during the initial session.

Mode The most frequent score in a distribution.

Modification The changes being made that influence the construct being measured, and, therefore, the scores do not retain the same meaning as the original instrument.

Multitrait-multimethod matrix A matrix that includes information on correlations between the measure and traits that it should be related to and traits that it should not theoretically be related to. The matrix also includes correlations between the measure of interest and other same-methods measures and measures that use different assessment methods.

Narrative recordings Recordings that typically do not involve any quantitative recording procedures. They can be completed by the counselor, parents, other family members, teachers, or the client.

Naturalistic observation A type of observation in which the counselor observes the client in a typical environment and the counselor does not manipulate any aspect of the environment during the observation.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Negatively skewed distribution A type of distribution where the majority of scores are on the higher end of the distribution.

Neuropsychological assessment An assessment of cognitive impairments, specifically the behavioral effects of possible brain injury or damage.

Nominal scale A scale of measurement characterized by assigning numbers to name or representing mutually exclusive groups (e.g., 1 = male, 2 = female).

Normal curve A bell-shaped, symmetrical, and unimodal curve. The majority of cases are concentrated close to the mean, with 68% of the individual scores falling between one standard deviation below the mean and one standard deviation above the mean.

Normal distribution A distribution of scores with certain specific characteristics (e.g., 68% of the sample approximately falls between one standard deviation below the mean to one standard deviation above the mean).

Norm-referenced instruments Instruments in which the interpretation of performance is based on the comparison of an individual's performance with that of a specified group of people.

Observation The most common method counselors use to assess personality in which they observe clients from the first meeting and begin to make clinical judgments based on those initial observations.

Ordinal scale Type of measurement scale in which the degree of magnitude is indicated by the rank ordering of the data.

Panic attack An abrupt surge of fear or intense discomfort that reaches its peak quickly, and clients experience symptoms such as palpitations, sweating, trembling, choking, chest pain, nausea, and fear of losing control.

Percentile rank A ranking that provides an indication of the percent of scores that fall at or below a given score. For example: "Mary's percentile of 68 means that if there were 100 people who had taken this instrument, 68 of them would have a score at or below Mary's."

Percentile scores Another term for *percentile ranks*.

Performance assessments An alternate method of assessing individuals, other than through multiple-choice types of items, in which the focus is on evaluating the performance of tasks or activities.

Performance tests Tests that require the manipulation of objects with minimal verbal influences.

Positively skewed distribution A type of distribution in which the majority of scores are at the lower end of the range of scores.

Predictive validity A type of validation evidence in which there is a delay between the time the instrument is administered and the time the criterion information is gathered.

Projective techniques A type of personality assessment that provides the client with a relatively ambiguous stimulus, thus encouraging a nonstructured response. The assumption underlying these techniques is that the individual will project his or her personality into the response. The interpretation of projective techniques is subjective and requires extensive training in the technique.

Psychological report A summary of a client's assessment results that is often geared toward other professionals. Frequently written by a psychologist, a typical report includes background information, behavioral observations, test results and interpretations, recommendations, and a summary.

Psychological test An objective and standardized measure of a sample of behavior.

Psychosocial interview A detailed interview that gathers background information and information about the client's current psychological and social situation.

Qualitative data Descriptive information sought from the evaluation study, with the intent to produce "rich" interpretative data.

Quantitative data Information that is more numerical in nature where the intent is to quantify the results.

Randomized clinical trials design The gold standard in intervention research. Adopted from the medical model, where patients are *randomly* assigned to receive the medication or the placebo, in counseling evaluation studies, clients are *randomly* assigned to either the intervention group or the placebo/control group.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Ratings recording A category of behavioral assessment in that rating scales are completed, but often by either the counselor or some other observer (e.g., parents or teachers).

Ratio scale A scale of measurement that has both interval data and a meaningful zero (e.g., weight, height). Because ratio scales have a meaningful zero, ratio interpretations can be made.

Raw scores Raw scores are the unadjusted scores on an instrument before they are transformed into standard scores. An example of a raw score is the number of answers an individual gets correct on an achievement test.

Reactivity The possible change that may occur in clients' behaviors, thoughts, or performance that are a result of being observed, assessed, or evaluated.

Regression A commonly used statistical technique in which the researcher examines whether independent variables predict to a criterion or dependent variable. Regression is used to determine if there is a linear relationship among the variables or a line of best fit.

Regression equation An equation that describes the linear relationship between the predictor variable(s) and the criterion variable. These equations are often used to determine if it is possible to predict the criterion based on the instrument's scores.

Reliability Concerns the degree to which a measure or a score is free of random error. In classical test theory, it is the ratio of true variance to observed variance.

Reliability generalization A meta-analytic method that combines estimates of reliability across studies in order to calculate an estimate based on multiple indicators of reliability.

Response-to-intervention The replacement for the discrepancy approach of diagnosing a learning disability, the focus of which is to use data (e.g., achievement tests, classroom activities) to identify students at risk for poor learning outcomes.

Score A number or letter that is the product of a client taking an assessment. A score cannot be interpreted without additional information about the assessment.

Self-monitoring The practice of observing and recording one's own behavior.

Semistructured interview An interview that is a combination of a structured and unstructured format in which there are a set of established questions and the clinician can also ask additional questions for elaboration or to gather additional information.

Sequential processing The use of mental abilities to arrange stimuli in sequential, or serial order in order to process the information.

Simple random sample A type of sample in which every individual in the population has an equal chance of being selected.

Simultaneous processing The use of mental abilities to integrate information in a unified manner, with the individual integrating fragments of information in order to comprehend the whole.

Skewed The distribution is not symmetrical and the majority of people either scored in the low range or the high range, as compared with a normal distribution in which the majority scored in the middle.

Skewed distributions Distributions in which the majority of scores are either high or low. Skewed distributions are asymmetrical, and the mean, mode, and median are different. In *positively skewed distributions,* the majority of scores are on the lower end of the distribution; in *negatively skewed distributions,* the majority of scores are on the upper end of the distribution.

Slope bias A term referring to a situation in which a test yields significantly different validity coefficients for different groups, resulting in different regression lines.

Spearman-Brown formula A formula for correcting a split-half reliability coefficient that estimates what the coefficient would be if the original number of items were used.

Spearman's model A two-factor theory of intelligence that postulates everyone has a *general ability factor* influencing their performance on intellectual tasks, and also specific factors correlated to *g* that influence performance in specific areas.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Split-half reliability One of the internal consistency measures of reliability in which the instrument is administered once and then split into two halves. The scores on the two halves are then correlated to provide an estimate of reliability. Often the split-half reliability coefficients are corrected using the Spearman-Brown formula. This formula adjusts the coefficient for using only half of the total number of items to provide an estimate of what the correlation coefficient would be if the original number of items was used.

Standard deviation The most common statistic used to describe the variability of a set of measurements. It is the square root of the variance.

Standard error of difference A measure used by a counselor to examine the difference between two scores and determine if there is a significant difference.

Standard error of estimate A numerical result that indicates the margin of expected error in the individual's predicted criterion score as a result of imperfect validity.

Standard error of measurement This deviation provides an indication of what an individual's true score would be if he or she took the instrument repeated times. Counselors can use standard error of measurement to determine the range of scores 68%, 95%, or 99.5% of the time.

Stratified sample A type of sample in which individuals are selected for the norming group based on certain demographic characteristics.

Structured interview An interview that is conducted using a predetermined set of questions that is asked in the same manner and sequence for every client.

Structured personality instruments Formalized assessments in which clients respond to a fixed set of questions or items.

Summative evaluation A cumulative evaluation of services that are typically completed at the endpoint of the service. These types of evaluation are designed to provide an overall indication of the effectiveness of the services.

Target behavior The behavior that is being assessed in behavior assessment.

Test An individual instrument in which the focus is on evaluation.

Testing A process of giving clients tests and/or instruments.

Test-retest reliability One in which the reliability coefficient is obtained by correlating a group's performance on the first administration of an instrument with the same group's performance on the second administration of that same instrument.

Test sophistication A term applied to an individual's level of knowledge in test-taking skills. It is not related to knowledge of the content but rather to the format of the tests and the skills required for maneuvering through that format.

Universal design An approach to instrument design with the goal of maximizing accessibility for all intended examinees.

Unstructured interview An interview in which the clinician gears the questions toward each individual client and there is no established set of questions.

Validity coefficient The correlation between the scores on an instrument and the criterion measure.

Validity generalization Term applied to findings indicating that the validity of cognitive ability tests can be generalized and that cognitive ability is highly related to job performance.

Variance The average of the squared deviation from the mean. It is a measure of variability and its square root is the standard deviation of the set of measurements.

Z scores A standard score that always has a mean of 0 and a standard deviation of 1.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

References

Abedi, J. (2013). Testing of English language learner students. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology, Volume 3* (pp. 355–368). Washington, DC: American Psychological Association.

Abedi, J., Hofstetter, C. A., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*, 1–28. doi:10.1207/s15326977ea1003_3

Ackerman, S. J., Hilsenroth, M. J., Baity, M. R., & Blagys, M. D. (2000). Interaction of therapeutic process and alliance during psychological assessment. *Journal of Personality Assessment, 75*, 82–109.

ACT. (2014). *Using your ACT results 2014–2015*. Retrieved from http://www.act.org/aap/pdf/Using-Your-ACT-Results.pdf

Adesso, V. J., Cisler, R. A., Larus, B. J., & Haynes, B. B. (2004). Substance abuse. In M. Herson (Ed.), *Psychological assessment in clinical practice* (pp. 147–174). New York: Brunner-Routledge.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341–382. doi: 10.1177/0011000005285875

Aiken, L. R., & Groth-Marnat, G. (2005). *Psychological testing and assessment* (12th ed.). Boston, MA: Allyn & Bacon.

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35*, 13–28. doi:10.1177/0022022103260382

Altman, J., Thurlow, M., & Vang, M. (2010). *Annual performance report: 2007–2008 state data*. Retrieved from http://www.cehd.umn.edu/NCEO/Online Pubs/APRreport2007-2008.pdf

American Counseling Association. (2003). *Standards for qualifications of test users*. Alexandria, VA: Author.

American Counseling Association. (2014). *ACA code of ethics*. Alexandria, VA: Author.

American Educational Research Association. (2000). *AERA position statement concerning high-stakes testing in pre-K–12 education*. Washington, DC: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Foundation for Suicide Prevention. (2014). *Facts and figures*. Retrieved from https://www.afsp.org/understanding-suicide/facts-and-figures

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental* (4th ed. text revision). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

American Psychological Association. (2001). *Appropriate use of high stakes testing in our nation's schools*. Washington, DC: Author.

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Washington, DC: Author.

American Psychological Association. (2013). *Guidelines for psychological practice with older adults*. Washington, DC: Author.

American School Counselor Association. (2010). *Ethical standards for school counselors*. Alexandria, VA: Author.

American School Counselor Association. (2012). *ASCA national model*. Alexandria, VA: Author.

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist*, *36*, 1086–1093.

Anastasi, A. (1993). A century of psychological testing: Origins, problems and progress. In T. K. Fagen & G. R. VandenBos (Eds.), *Exploring applied psychology: Origins and critical analyses* (pp. 3–36). Washington, DC: American Psychological Association.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Archer, R. P. (1992). Review of the Minnesota Multiphasic Personality Inventory 2. In J. C. Conoley & J. J. Kramer (Eds.), *Eleventh mental measurements yearbook* (pp. 558–562). Lincoln, NE: The University of Nebraska Press.

Archer, R. P., & Krishnamurthy, R. (1996). The Minnesota Multiphasic Personality Inventory—Adolescent. In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 59–107). Boston, MA: Allyn & Bacon.

Archer, R. P., & Newsom, C. R. (2000). Psychological test usage with adolescent clients: Survey update. *Assessment, 7*, 227–235.

Armstrong, P. I., Day, S. X., McVay, J. P., & Rounds, J. (2008). Holland's RIASEC model as an integrative framework for individual differences. *Journal of Counseling Psychology, 55*, 1–18.

Association for Assessment in Counseling. (2003). *Responsibilities of Users of Standardized Tests (RUST)* (3rd ed.). Alexandria, VA: Author.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Association for Assessment in Counseling and Education. (2012). *Standards for multicultural assessment* (4th ed.). Alexandria, VA: Author.

Athanasou, J. A., & Cooksey, R. W. (1993). Self-estimates of vocational interests. *Australian Psychologist, 28*, 118–127.

Austin, J. T. (1994). Test review: Minnesota Multiphasic Personality Inventory 2. *Measurement and Evaluation in Counseling and Development, 27*, 178–185.

Baghurst, P. A., McMichael, A. J., Wigg, N. R., Vimpani, G. V., Robertson, E. F., Roberts, R. J., et al. (1992). Environmental exposure to lead and children's intelligence at the age of seven years: The Port Pirie cohort study. *New England Journal of Medicine, 327*, 1279–1284.

Baker, F. B. (2001). *The basics of item response theory*. College Park, MD: ERIC.

Baldwin, A. L., Kalhorn, J., & Breese, F. H. (1945). Patterns of parent behavior. *Psychological Monographs, 58* (Whole No. 268).

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.

Banks, S. R. (2012). *Classroom assessment: Issues and practices* (2nd ed.). Long Grove, IL: Waveland.

Barak, A. (2003). Ethical and professional issues in career assessment on the Internet. *Journal of Career Assessment, 11*, 3–21.

Barak, A., & English, N. (2002). Prospects and limitations of psychological testing on the Internet. *Journal of Technology in Human Services, 19*, 65–89.

Barthlow, D. L., Graham, J. R., Ben-Porath, Y. S., & McNulty, J. L. (1999). Incremental validity of the MMPI-2 Content Scales in an outpatient mental health setting. *Professional Assessment, 11*, 39–47.

Bayley, N., & Oden, M. H. (1955). The maintenance of intellectual ability in gifted adults. *Journal of Gerontology, 10*, 91–107.

Beck, A. T., & Steer, R. A. (1991). *Beck Scale for Suicide Ideation*. San Antonio, TX: Psychological Corporation.

Beck, A. T., & Steer, R. A. (1993). *Beck Hopelessness Scale*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory II*. San Antonio, TX: Psychological Corporation.

Becker, K. A. (2003). History of the Stanford-Binet intelligence scales: Content and psychometrics. (Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 1). Itasca, IL: Riverside Publishing.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1990). *Differential Aptitude Tests* (5th ed.). San Antonio, TX: Psychological Corporation.

Ben-Porath, Y. F., & Tellegen, A. (2008). *Minnesota Multiphasic Personality Interventory-2 Restructured Form: User's guide for reports*. Minneapolis: University of Minnesota Press.

Benziman, H., & Toder, A. (1993). The psychodiagnostic experience: Reports of subjects. In B. Nevo & R. S. Jager (Eds.), *Educational and psychological testing: The test taker's outlook* (pp. 287–299). Toronto: Hogrefe & Huber.

Bernt, F. (2001). Review of the Marital Satisfaction Inventory—Revised. In B. Plake & J. Impara (Eds.), *The fourteenth mental measurements yearbook* (pp. 711–712). Lincoln, NE: Buros Institute of Mental Measures.

Bertone, A., Bettinelli, L., & Faubert, J. (2007). The impact of blurred vision on cognitive assessment. *Journal of Clinical and Experimental Neuro-psychology, 29*, 467–476. doi: 10.1080/13803390600770793

Betsworth, D. G., & Fouad, N. A. (1997). Vocational interests: A look at the past 70 years and a glance at the future. *The Career Development Quarterly, 46*, 23–41.

Betz, N. E. (1988). The assessment of career development and maturity. In W. B. Walsh & S. H. Osipow (Eds.), *Career decision making* (pp. 77–136). Hillsdale, NJ: Erlbaum.

Betz, N. E. (1992). Career assessment: A review of critical issues. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (2nd ed., pp. 453–484). New York: Wiley.

Betz, N. E. (2005). Women's career development. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 253–277). Hoboken, NJ: Wiley.

Betz, N. E., Borgen, F. H., & Harmon, L. W. (2005). *Skills Confidence Inventory: Research, development, and strategies for interpretation*. Mountain View CA: CPP, Inc.

Betz, N. E., Klein, K. L., & Taylor, K. M. (1996). Evaluation of a short form of the Career Decision-Making Self-Efficacy Scale. *Journal of Career Assessment*, *4*, 47–57. doi:10.1177/106907279600400103

Betz, N. E., & Taylor, K. M. (1994). *Manual for the Career Decision-Making Self-Efficacy Scale*. Columbus, OH: Ohio State University Department of Psychology.

Beutler, L. F., Malik, M., Talebi, H., Fleming, J., & Moleiro, C. (2004). Use of psychological tests/ instruments for treatment planning. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (vol. 1, 3rd ed., pp. 111–145). Mahwah, NJ: Erlbaum.

Bickman, L., Rosof-Williams, J., Salzer, M. S., Summerfelt, W. T., Noser, K., Wilson, S. J., & Karver, M. S. (2000). What information do clinicians value for monitoring adolescent client progress and outcomes? *Professional Psychology: Research and Practice, 31*, 70–74.

Bischof, G. H., & Helmeke, K. B. (2005). Family sculpture procedures. In C. Manfred, T. Volker, & D. H. Sprenkle (Eds.), *Family assessment: Integrating multiple perspectives* (pp. 257–281). Ashland, OH: Hogrefe & Huber.

Bjorklund, D. F. (2005). *Children's thinking: Developmental function and individual differences* (4th ed.). Belmont, CA: Brooks/Cole.

Bjorklund, D. F. (2012). *Children's thinking: Cognitive development and individual differences* (5th ed.). Belmont, CA: Wadsworth/Cengage.

Blacher, J. H., Murray-Ward, M., & Uellendahl, G. E. (2005). School counselors and student assessment. *Professional School Counseling, 8*, 337–343.

Bond, L. (1989). The effects of special preparation on measures of scholastic ability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 429–444). New York: Macmillan.

Bornstein, R. F., Rossner, S. C., Hill, E. L., & Stephanian, M. L. (1994). Face validity and fakability of objective and projective measures of dependency. *Journal of Personality Assessment, 63*, 363–386.

Boughner, S. R., Hayes, S. R., Bubenzer, D. L., & West, J. D. (1994). Use of standardized assessment instruments by marital and family therapists: A survey. *Journal of Marital and Family Therapy, 20*, 69–75.

Bowen, M. (1978). *Family therapy in clinical practice*. New York: Jason Aronson.

Bowman, M. L. (1989). Testing individual differences in ancient China. *American Psychologist, 41*, 1059–1068. doi: 10.1037/0003-066X.44.3.576.b

Boyle, G. J. (1995). Review of the Rotter Incomplete Sentences Blank. In J. C. Conoley & J. J. Kramer (Eds.), *Twelfth mental measurements yearbook* (pp. 880–882). Lincoln, NE: The University of Nebraska Press.

Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating, and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12*, 119–132.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Braden, J. P., & Ouzts, S. M. (2005). Review of the Kaufman Assessment Battery for Children. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (2nd ed., pp. 517–520). Lincoln, NE: Buros Institute of Mental Measurements.

Bradley-Johnson, S., & Ekstrom, R. (1998). Visual impairments. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. D. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 271–295). Washington, DC: American Psychological Association.

Branch, L. G., & Jette, A. M. (1982). A prospective study of long-term care institutionalization among the aged. *American Journal of Public Health, 72*, 1373–1379.

Branch, L. G., Jette, A. M., & Evashwick, C. (1981). Toward understanding elders' health service utilization. *Journal of Community Health, 7*, 80–92.

Brauer, B. A., Braden, J. P., Pollard, R. Q., & Hardy-Braz, S. T. (1998). Deaf and hard of hearing people. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. D. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 297–315). Washington, DC: American Psychological Association.

Bray, D. W. (1982). The assessment center and the study of lives. *American Psychologist, 37*, 180–189. doi: 10.1037/0003-066X.37.2.180

Bray, J. H. & Stanton, M. (2009). Couple and family assessment. In *The handbook of family psychology* (pp. 151–163). Oxford: Wiley-Blackwell.

Brooke, S. L., & Ciechalski, J. C. (1994). Minnesota Importance Questionnaire. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (pp. 222–225). Alexandria, VA: National Career Development Association.

Brown, G. S., & Lambert, M. J. (1998, June). Tracking patient progress: Decision making for cases who are not benefiting from therapy. Paper presented at the 29th annual meeting of The Society for Psychotherapy Research, Snowbird, UT.

Brown, J., Dreis, S., & Nace, D. K. (1999). What reallymakes a difference in psychotherapy outcome? Why does managed care want to know? In M. A. Hubble, B. L. Duncan, & S. D. Miller (Eds.), *The heart & soul of change: What works in therapy* (pp. 389–406). Washington, DC: American Psychological Association.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (2010). *Test of Nonverbal Intelligence* (4th ed.). San Antonio, TX: Pearson.

Brown, S. D., Lent, R. W., & Gore, P. A. (2000). Self-rated abilities and self-efficacy beliefs: Are they empirically distinct? *Journal of Career Assessment, 8*, 223–235. doi: 10.1177/106907270000800302

Bubany, S. T., & Hanson, J. C. (2010). Ability self-estimates and self-efficacy: Meaningfully distinct? *Measurement and Evaluation in Counseling and Development, 43*(3), 168–187. doi: 10.1177/0748175610384809

Bubenzer, D. L., Zimpfer, D. G., & Mahrle, C. L. (1990). Standardized individual appraisal in agency and private practice: A survey. *Journal of Mental Health Counseling, 12*, 51–66.

Buck, J. (1948). The H-T-P. *Journal of Clinical Psychology, 4*, 151–159.

Buck, J. (1992). *The House-Tree-Person projective drawing technique: Manual and interpretive guide* (Rev. ed.). Los Angeles, CA: Western Psychological Press.

Bujold, C. (2004). Constructing career through narrative. *Journal of Vocational Behavior, 64*, 470–484.

Burlingame G. M., Cox, J. C., Wells, M. G., Lambert, M. J., Latkowski, M., et al. (2005). *The Administration and Scoring Manual of the Youth Outcome Questionnaire*. Salt Lake City, UT: American Professional Credential Service.

Burns, R. (1982). *Self-growth in families: Kinetic Family Drawings (K-F-D) research and application*. New York: Brunner/Mazel.

Burns, R. (1987). *Kinetic-House-Tree-Person Drawings (K-H-T-P)*. New York: Brunner/Mazel.

Burns, R., & Kaufman, S. (1970). *Kinetic Family Drawings (K-F-D): An introduction to understanding children through kinetic drawings*. New York: Brunner/Mazel.

Burns, R., & Kaufman, S. (1972). *Action, styles, symbols in Kinetic Family Drawings (K-F-D)*. New York: Brunner/Mazel.

Burton, N. W., & Wang, M. (2005). *Predicting longterm success in graduate school: A collaborative validity study*. GRE Board Report No. 99-14R. Retrieved from http://www.ets.org/Media/Research/pdf/ RR-05-03.pdf

Busby, D. M., Holman, T. B., & Taniguchi, N. (2001). RELATE: Relationship evaluation of the individual, family, cultural, and couple contexts. *Family Relations, 50*, 308–316.

Butcher, J. N. (1995). Clinical personality assessment: An overview. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 3–9). New York: Oxford University Press.

Butcher, J. N. (2011). *A beginner's guide to the MMPI-2* (3rd ed.). Washington, DC: American Psychological Association.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kraemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kraemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring* (rev. ed.). Minneapolis: University of Minnesota Press.

Butcher, J. N., & Williams, C. L. (2000). *Essentials of the MMPI-2 and MMPI-A interpretation* (2nd ed.). Minneapolis: University of Minnesota Press.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R., Tellegen, A., Ben-Porath, Y. S., & Kraemmer, B. (1992). *MMPI-A manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.

Butcher, J. N. (1995). Clinical personality assessment: An overview. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 3–9). New York: Oxford University Press.

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31*, 141–154. doi: 10.1037/0735-7028.31.2.141

Campbell, D. P. (1992). *Campbell Interest and Skills Survey manual*. Minneapolis: National Computer Systems.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multi-trait multi-method matrix. *Psychological Bulletin, 56*, 81–105.

Campbell, J. P. (1994). Alternative models of job performance and their implications for selection and classification. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 33–51). Hillsdale, NJ: Erlbaum.

Campbell, V. L. (2000). A framework for using tests in counseling. In C. E. Watkins & V. L. Campbell (Eds.), *Testing and assessment in counseling practice* (2nd ed., pp. 3–11). Mahwah, NJ: Erlbaum.

Caprara, G. V., Barbaranelli, C., & Comfrey, A. L. (1995). Factor analysis of the NEO-PI Inventory and the Comfrey Personality

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Scales in an Italian sample. *Personality and Individual Differences, 18*, 193–200.

Capraro, R. M., & Capraro, M. M. (2002). Myers-Briggs Type Indicator score reliability across studies: A meta-analytic reliability generalization study. *Educational and Psychological Measurement, 62*(4), 590–602. doi: 10.1177/0013164402062004004

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122–130). New York: Guilford.

Carter, E. A., McGoldrick, M., & Garcia-Preto, N. (2010). *The expanded family life cycle: Individual, family, and social perspectives* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Cattell, R. B. (1979). Are culture fair intelligence tests possible and necessary? *Journal of Research and Development in Education, 12*(2), 3–13.

Cattell, R. B., Cattell, A. K., & Cattell, H. E. (1993). *Sixteen Personality Factor Questionnaire* (5th ed.). Champaign, IL: Institute for Personality and Ability Testing.

Cawthon, S. W. (2009). Accommodations for students who are deaf or hard of hearing in large-scale, standardized assessments: Surveying the landscape and charting a new direction. *Education Measurement: Issues and Practice, 28*(2), 41–49. doi: 10.1111/j.1745-3992.2009.00147.x

Ceci, S. J. (1990). *On intelligence—More or less: A bioecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice Hall.

Ceci, S. J. (1991). How much does school influence general intelligence and its cognitive components: A reassessment of the evidence. *Developmental Psychology, 27*, 703–723.

Ceci, S. J. (1993). Contextual trends in intellectual development. *Developmental Review, 13*, 403–435.

Center for Disease Control and Prevention. (2011). CDC report: Mental illness surveillance among U.S. adults. Retrieved from http://www.cdc.gov/mentalhealthsurveillance/

Center for Disease Control and Prevention (2015). *Suicide: Facts at a glance.* Retrieved from http://www.cdc.gov/ViolencePrevention/pdf/Suicide-DataSheet-a.pdf

Chang, V. N., Scott, S. T., & Decker, C. L. (2012). *Developing helping skills: A step-by-step approach to competency* (2nd ed.). Belmont, CA: Brooks/Cole Cengage.

Chartrand, J. M., & Walsh, W. B. (2001). Career assessment: Changes and trends. In F. T. L. Leong & A. Barak (Eds.), *Contemporary models in vocational psychology* (pp. 231–256). Mahwah, NJ: Lawrence Erlbaum.

Child Welfare Information Gateway. (2013). *What is child abuse and neglect? Recognizing the signs and symptoms*. Washington, DC: U.S. Department of Health and Human Services, Children's Bureau.

Childstats.gov. (2013). *American's children: Key national indicators of well-being*. Retrieved from http://www.childstats.gov/americaschildren13/famsoc5.asp

Chipuser, H. M., Rovine, M., & Plomin, R. (1990). LISREL modeling: Genetic and environmental influences on IQ revisited. *Intelligence, 14*, 11–29.

Chittooran, M. M. (2010). Review of Wechsler Memory Scale. In K. F. Geisinger, R. A. Spies, & J. F. Carlson. *The eighteenth mental measurement yearbook* (4th ed.). [electronic version]

Chope, R. C. (2015). Card sorts, sentence completions, and other qualitative assessments. In P. J. Hartung, M. L. Savickas, & W. B. Walsh (Eds.), *APA handbook of career interventions. Vol. 2 applications* (pp. 7184). Washington, DC: American Psychological Association.

Ciechalski, J. C. (2009). Self-Directed Search. In E. A. Whitfield, R. W. Feller, & C. Wood (Eds.), *A counselor's guide to career assessment instruments* (5th ed., pp. 296–308). Broken Arrow, OK: National Career Development Association.

Cierpka, M. (2005a). Introduction to family assessment. In M. Cierpka, V. Thomas, & D. H. Sprenkle (Eds.), *Family assessment: Integrating multiple perspectives* (pp. 3–14). Cambridge, MA: Hogrefe.

Cierpka, M. (2005b). The three-level model of family assessment. In M. Cierpka, V. Thomas, & D. H. Sprenkle (Eds.), *Family assessment: Integrating multiple perspectives* (pp. 15–32). Cambridge, MA: Hogrefe.

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and emerging reconceptualization. *Annual Review of Psychology, 58*, 227–257. doi: 10.1146/annurev.psych.57.102904. 190200

Clarkin, J. F., & Levy, K. N. (2004). The influence of client variables on psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 227–306). New York: Wiley.

Cohen, L., Duberly, J., & Mallon, M. (2004). Social constructionism in the study of career: Accessing the parts that other approaches cannot reach. *Journal of Vocational Behavior*, *64*, 979–1006.

Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2013). *Psychological testing and assessment: An introduction to tests and measurement* (8th ed.). New York: McGraw-Hill.

Conners, C. K. (2008). *Conners* (3rd ed.). North Tonawanda, NY: Multi-Health Systems.

Cook, W. L., & Kenny, D. A. (2004). Application of the social relations model to family assessment. *Journal of Family Psychology, 18*, 361–371.

Cormier, S., Nurius, P. S., & Osborn, C. J. (2009). *Interviewing and change strategies for helpers: Fundamental skills and cognitive behavioral interventions* (6th ed.). Belmont, CA: Brooks/Cole.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Council for the Accreditation of Counseling and Related Educational Programs. (2009). *The 2009 CACREP standards*. Alexandria, VA: Author.

Coupland, S. K., Serovic, J., & Glenn, J. E. (1995). Reliability in constructing genograms: A study among marriage and family therapy doctoral students. *Journal of Marital and Family Therapy, 21*, 251–263.

Craig, R. J. (1993). *The Millon Clinical Multiaxial Inventory: A clinical research information synthesis*. Hillsdale, NJ: Erlbaum.

Crites, J. O. (1973). *The Career Maturity Inventory.* Monterey, CA: CTB/McGraw-Hill.

Crites, J. O. (1978). *The Career Maturity Inventory* (2nd ed.). Monterey, CA: CTB/McGraw-Hill.

Crites, J. O., & Savickas, M. L. (1995). *Career Maturity Inventory: Sourcebook*. Oroville, WA: Bridges.com.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace Jovanovich.

Cromwell, R., Fournier, D., & Kvebaek, D. (1980). *The Kvebaek Family Sculpture Technique: A diagnostic and research tool in family therapy*. Jonesboro, TN: Pilgrimage.

Cronbach, L. J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika, 16*, 297–334.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: HarperCollins.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Champaign, IL: University of Illinois Press.

Cronbach, L. J., Gleser, G. C., Rajaratnam, N., & Nanda, H. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Cull, J. G., & Gill, W. S. (1992). *Suicide Probability Scale*. Los Angeles: Western Psychological Services.

Daily, S. F., Gill, C. S., Karl, S. L., & Barrio Minton, C. A. (2014). *DSM-5 learning companion for counselors.* Alexandria, VA: American Counseling Association.

Dana, R. H. (2005). *Multicultural assessment: Principles, applications, and examples*. Mahwah, NJ: Erlbaum.

Daniels, M. H. (1994). Self-Directed Search. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (pp. 206–212). Alexandria, VA: National Career Development Association.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myths*. New York: Free Press.

Day, S. X., & Rounds, J. (1998). Universality of vocational interest structure among racial and ethnic minorities. *American Psychologist, 53*(7), 728736. doi: 10.1037/0003-066X.53.7.728

DeBell, C. (2006). What all applied psychologists should know about work. *Professional Psychology: Research and Practice, 37*(4), 325–333. doi: 10.1037/0735-7028.37.4.325

Deisinger, G., Randazzo, M., O'Neill, D., & Savage, J. (2008). *The handbook for campus threat assessment and management teams.* Boston, MA: Applied Risk Management.

Derogatis, L. R. (1993). *The Brief Symptom Inventory (BSI) administration, scoring, and procedures*. Minneapolis: Pearson Assessments.

Derogatis, L. R. (1994). *Administration, scoring, and procedures manual for the SCL-90-R* (3rd ed.). Minneapolis: Pearson Assessments.

Derogatis, L. R. (2000). *Brief Symptoms Inventory-18*. Minnetonka, MN: NCS Assessments.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*(1), 7175.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*, 417–440.

Digman, J. M., & Takemoto-Chock, N. K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research, 16*, 149–170.

Dimitrov, D. (2007). Reliability. In B. T. Erford (Ed.), *Assessment for counselors* (pp. 99–122). Boston, MA: Lahaska Press.

Dollarhide, C. T., & Lemberger, M. E. (2006). No child left behind: implications for school counselors. *Professional School Counseling, 9*, 295–304.

Dominque, B. W., & Briggs, D. C. (2009). Using linear regression and propensity score matching to estimate the effect of coaching on the SAT. *Multiple Linear Regression Viewpoints, 35*, 12–29.

Donlon, T. F. (Ed.). (1984). *The College Board technical handbook for the Scholastic Aptitude Test and achievement testing*. New York: College Board Publications.

Donnay, D. A. C., Morris, M. L., Schaubhut, N. A., & Thompson, R. C. (2005). *Strong Interest Inventory manual*. Mountain View, CA: Consulting Psychologists Press.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory—II. *Psychological Assessment*, *10*, 83–89.

Duffy, M., & Chenail, R. J. (2004). Qualitative strategies in couple and family assessment. In L. Sperry (Ed.), *Assessment of couples and families: Cutting-edge strategies* (pp. 33–64). New York: Brunner-Routledge.

Duffy, M., & Chenail, R. J. (2012). Qualitative assessment. In L. Sperry (Ed.), *Assessment of couples and families: Contemporary and cutting-edge strategies* (2nd ed,. pp. 17–51). New York: Routledge.

Duley, S. M., Cancelli, A. A., Kratochwill, T. R., Bergan, J. R., & Meredith, K. E. (1983). Training and generalization of motivational analysis interview assessment skills. *Behavioral Assessment, 5*, 281–293.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth edition* (4th ed.). Bloomington, MN: Pearson.

Duran, R. P. (1983). *Hispanics' education and background: Predictors of college achievement*. New York: College Entrance Examination Board.

Ekstrom, R. B., Elmore, P. B., Schafer, W. D., Trotter, T. V., & Webster, B. (2004). A survey assessment and evaluation activities of school counselors. *Professional School Counseling, 8*, 24–30.

Elliot, C. D. (2007). *Differential Ability Scale-Second Edition: Technical manual*. San Antonio, TX: Pearson.

Elmore, P. B., Ekstrom, R., Diamond, E. E., & Whittaker, S. (1993). School counselors' test use patterns and practices. *The School Counselor, 41*, 73–80.

Embretson, S. E. (2010). Measuring psychological constructs with model-based approaches: An introduction. In S. E. Embretson (Ed.), *Measuring psychological constructs* (pp. 1–7). Washington, DC: American Psychological Association.

Embretson, S. E., & Hershberger, S. L. (1999). *The new rules of measurement: What every psychologist and educator should know*. Mahwah, NJ: Erlbaum.

Epperson, D. L., Bushway, D. J., & Warman, R. E. (1983). Client self-termination after one session: Effects of problem recognition, counselor gender, and counselor experience. *Journal of Counseling Psychology, 30*, 307–315.

Eriksen, K., & Kress, V. E. (2006). The DSM and the professional counseling identity: Bridging the gap. *Journal of Mental Health Counseling, 28*(3), 202217.

Erford, G. T., Klein, L., & McNinch, K. (2007). Assessment of intelligence. In B. T. Erford (Ed.), *Assessment for counselors* (pp. 45–98). Boston, MA: Lahaska.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*, 246–262.

Evans, K. M. (2013). Multicultural considerations in career assessment. In C. Wood & D. G. Hayes (Eds.), *A counselor's guide to career assessment instruments* (6th ed., pp. 4961). Broken Arrow, OK: National Career Development Association.

Eyde, L. D., Robertson, G. J., & Krug, S. E. (2010). *Responsible test use: Case studies for assessing human behavior* (2nd ed.). Washington, DC: American Psychological Association.

Fernandez, K., Boccaccini, M. T., & Noland, R. M. (2007). Professionally responsible test selection for Spanish-speaking clients: A four-step approach for identifying and selecting translated tests. *Professional Psychology: Research and Practice, 38*, 363–374.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Finn, S. E. (2007). *In our clients' shoes: Theory and techniques of therapeutic assessment*. New York: Taylor & Francis Group.

Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment, 4*, 278–287.

Fisher, L., & Sorenson, G. P. (1996). *School law for counselors, psychologists, and social workers*. White Plains, NY: Longman.

Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert & S. T. Fiske (Eds.), *Handbook of social psychology: Vol. 2* (4th ed., pp. 915–981). New York: McGraw-Hill.

Fitts, W. H., & Warren, W. L. (1997). *Tennessee Self-Concept Scale: Second Edition (TSCS:2)*. Los Angeles: Western Psychological Services.

Flanagan, R. (2005). Review of the Achenbach System of Empirically Based Assessment. In Spies, R. A., & Plake, B. S. (Eds.). *The sixteenth mental measurements yearbook* (pp. 10–13). Lincoln, NE: Buros Institute of Mental Measurements.

Fleenor, J. W. (2001). Review of the Myers-Briggs Type Indicator. In B. Plake & J. Impara (Eds.), *The fourteenth mental measurements yearbook* (pp. 816–818). Lincoln, NE: Buros Institute of Mental Measures.

Flores, L. Y., Spanierman, L. B., & Obasi, E. M. (2003). Ethical and professional issues in career assessment with diverse racial and ethnic groups. *Journal of Career Assessment*, *11*, 76–95.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29–51. doi: 10.1037/0033-2929.95.129

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191. doi: 10.1037/h0090408

Flynn, J. R. (2013). The "Flynn Effect" and Flynn's paradox. *Intelligence, 41*, 851857. doi: 10.1016/j.intell.2013.06.0

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Fong, M. L. (1995). Assessment and *DSM-IV* diagnosis of personality disorders: A primer for counselors. *Journal of Counseling and Development, 73*, 635–639.

Fortier, L. M., & Wanlass, R. L. (1984). Family crisis following the diagnosis of a handicapped child. *Family Relations, 33*, 13–24.

Fouad, N. A., & Byars-Winston, A. M. (2005). Cultural context of career choice: Meta-analysis of race/ethnicity differences. *Career Development Quarterly, 53*, 223–233.

Fouad, N. A., Harmon, L. W., & Borgen, F. H. (1997). Structure of interests in employed male and female members of US racial-ethnic minority and nonminority groups. *Journal of Counseling Psychology, 44*(4), 339345. doi: 10.1037/0022-0167.44.4.339

Frank, R. G., Rosenthal, M., & Caplan, B. (2009). *Handbook of rehabilitation psychology* (2nd ed.). Washington, DC: American Psychological Association.

Fredman, N., & Sherman, R. (1987). *Handbook of measurements for marriage and family therapy*. New York: Brunner/Mazel.

Frisch, M. B. (1994). *Quality of Life Inventory*. Minneapolis: Pearson Assessments.

Futris, T. G., Barton, A. W., Aholou, T. M., & Seponski, D. M. (2011). The impact of PREPARE on engaged couples: Variations by delivery format. *Journal of Couple & Relationship Therapy, 10*, 6986.

Garber, A. M. (1989). Long-term care, welfare, and health of the disabled elderly living in the community. In David A. Wise (Ed.), *The economics of aging*. (pp. 255–278). Chicago: University of Chicago Press.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences* (10th anniversary ed.). New York: Basic Books.

Gati, I., Krausz, M., & Osipow, S. H. (1996). A taxonomy of difficulties in career decision-making. *Journal of Counseling Psychology, 43*, 510–526.

Gati, I., Osipow, S. H., Krausz, M., & Saka, N. (2000). Validity of the Career Decision-Making Questionnaire: Counselee versus career counselor perceptions. *Journal of Vocational Behavior*, *56*, 99–113.

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist, 34*, 127–140.

Gersten, A. (2013). *Integrative assessment: A guide for counselors.* Upper Saddle River, NJ: Pearson.

Gibbons, M. M. (2013). Kuder Career Planning System, In C. Wood & D. G Hays (Eds.), *A counselor's guide to career assessment instruments* (6th ed., pp. 203–210). Broken Arrow, OK: National Career Development Association.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social science research*. Thousand Oaks, CA: Sage.

Glover, C. (2010). Motivational interviewing. In P. Phillips, O. McKeown, & T. Sanford (Eds.), *Dual diagnosis: Practice in context* (pp. 72–79). Hoboken, NJ: Wiley-Blackwell.

Glutting, J. J., & Wilkinson, G. S. (2003). *Wide Range Interest and Occupational Test* (2nd ed.). Austin, TX: PRO-ED Inc.

Goldberg, L. R. (1992). The development of markers of the Big-Five factor structure. *Psychological Assessment, 4*, 26–42.

Goldberg, L. R. (1994). The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.

Goldman, L. (1990). Qualitative assessment. *The Counseling Psychologist, 18*, 205–213.

Goodyear, R. K. (1990). Research on the effects of test interpretation: A review. *The Counseling Psychologist, 18*, 240–257.

Gordon, H. W., & Lee, P. (1986). A relationship between gonadotroins and visuospatial function. *Neuropsycholia, 24*, 563–576.

Gottfredson, L. S. (1994). The science and politics of race norming. *American Psychologist, 49*, 955–963. doi: 10.1037/0003-066X.49.11.955

Gottfredson, L. S. (2009). Logical fallacies used to dismiss the evidence on intelligence testing. In R.P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 11–65). Washington, DC: American Psychological Association. doi: http://dx.doi.org/10.1037/11861-00

Gottfredson, L. S., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology, 50*, 183–195. doi: 10.1037/a0016641

Gough, H. G., & Bradley, P. (1996). *CPI manual* (3rd ed.). Palo Alto, CA: CPP, Inc.

Graduate Record Examination Board. (2014). *Graduate Record Examination: Guide to the use of scores 20142015*. Princeton, NJ: Educational Testing Services.

Graham, J. M., Liu, Y. J., & Jeziorski, J. L. (2006). The dyadic adjustment scale: A reliability generalization meta-analysis. *Journal of Marriage and Family, 68*(3), 701–717. doi: 10.1111/j.1741-3737.2006.00284.x

Graham, J. M., Liu, Y. J., & Jeziorski, J. L. (2006). The Dyadic Adjustment Scale: A reliability generalization meta-analysis. *Journal of Marriage and Family, 68*, 701–717.

Graham, J. R. (2006). *MMPI-2: Assessing personality and psychopathology* (4th ed.). New York: Oxford University Press.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Granello, D. G. (2010). The process of suicide risk assessment: Twelve core principles. *Journal of Counseling & Development*, *88*, 363–370.

Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., Dufour, M. C., & Pickering, R. P. (2004). The 12-month prevalence and trends in *DSM-IV* alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. *Drug and Alcohol Dependence, 74*, 223–234. doi: 10.1016/j.drugalcdep.2004.02.004

Greene, R. L., & Banken, J. A. (1995). Assessing alcohol/drug abuse problems. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 460–474). New York: Oxford University Press.

Greenspan, S. I., & Greenspan, N. T. (1981). *The clinical interview of the child*. New York: McGraw-Hill.

Gregory, R. J. (2011). *Psychological testing: History, principles, and applications* (6th ed.). Boston, MA: Allyn & Bacon.

Groth-Marnat, G. (2009a). *Handbook of psychological assessment* (5th ed.). Hoboken, NJ: Wiley.

Groth-Marnat, G. (2009b). The five assessment issues you meet when you go to heaven. *Journal of Personality Assessment, 91*(4), 303–310. doi: 10.1080/00223890902935662

Groth-Marnat, G., & Davis, A. (2013). *Psychological report writing assistance*. Hoboken, NJ: Wiley.

Guilford, J. B. (1988). Some changes in the structure of the intellect model. *Educational and Psychological Measurement, 48*, 1–4.

Hackett, G., & Lonborg, S. D. (1993). Career assessment for women: Trends and issues. *Journal of Career Assessment, 3*, 197–216.

Hackett, G., & Lonborg, S. D. (1994). Career assessment and counseling for women. In W. B. Walsh & S. H. Osipow (Eds.), *Career counseling for women* (pp. 43–86). Hillsdale, NJ: Erlbaum.

Hackett, G., & Watkins, C. E. (1995). Research in career assessment: Abilities, interests, decision making, and career development. In W. B. Walsh & S. H. Osipow (Eds.), *Handbook of vocational psychology: Theory, research, and practice* (2nd ed., pp. 181–215). Hillsdale, NJ: Erlbaum.

Hagen, J. W., & Hagen, W. W. (1995). What employment counselors need to know about employment discrimination and the Civil Rights Act of 1991. *Journal of Employment Counseling, 32*, 2–10.

Hampton, R. B., & Beavers, W. R. (2012). Observational assessment. In L. Sperry (Ed.), *Family assessment: Contemporary and cutting-edge strategies* (2nd ed., pp. 83114). New York: Routledge.

Handler, L. (1996). The clinical use of figure drawings. In C. S. Newmark (Ed.), *Major psychological assessment instruments* (2nd ed., pp. 206–293). Boston, MA: Allyn & Bacon.

Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of Clinical Psychology, 61*(2), 155–163. doi: 10.1002/jclp.20108

Hansen, J. C., & Leuty, M. E. (2007). Evidence of validity for the Skills Scale scores of the Campbell Interest and Skills Survey. *Journal of Vocational Behavior, 71*, 23–44.

Harmon, L. W. (1994). Career Decision Scale (CDS). In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (pp. 258–262). Alexandria, VA: National Career Development Association.

Harmon, L. W. (1997). Do gender differences necessitate separate career development theories and measures? *Journal of Career Assessment, 5*, 463–470.

Hartung, P. J. (1999). Interest assessment using card sorts. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Meaning, measurement, and counseling use* (pp. 235–252). Palo Alto, CA: Davies-Black.

Hattie, J. (1998). Review of the Tennessee Self-Concept Scale. In C. W. Conoley & J. Impara (Eds.), *The thirteenth mental measurements year-book* (2nd ed., pp. 1011–1012). Lincoln, NE: Buros Institute of Mental Measures.

Haynes, S., Follingstad, T., & Sullivan, J. (1979). Assessment of marital satisfaction and interaction. *Journal of Consulting and Clinical Psychology, 47*, 789–791.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Helms, J. E. (2002). A remedy for the Black-white test score disparity. *American Psychologist, 57*, 30–35.

Helms, J. E. (2004). Fair and valid use of educational testing in grades K-12. In J. E. Wall & G. R. Walz (Eds.), *Measuring up: Assessment issues for teachers, counselors, and administrators* (pp. 81–88). Greensboro, NC: CAPS Press.

Helms, J. E. (2006). Fairness is not validity or cultural bias in racial-group assessment: A quantitative perspective. *American Psychologist, 61*, 845–859.

Helms, J. E., & Ekstrom, R. B. (2002). Counseling assessment. In R. B. Ekstrom & D. K. Smith (Eds.), *Assessing individuals with disabilities in education, employment, and counseling settings* (pp.121–132). Washington, DC: American Psychological Association.

Helms, J. E., Jernigan, M., & Mascher, J. (2005). The meaning of race in psychology and how to change it. *American Psychologist, 60*, 27–36.

Henard, D. H. (2000). Item response theory. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 67–97). Washington, DC: American Psychological Association.

Henderson, K. L. (2013). Mandated reporting of child abuse: Considerations and guidelines for mental health counselors. *Journal of Mental Health Counseling, 35*(4), 296309.

Heppner, M. J. (1991). The Career Transitions Inventory (Available from M. J. Heppner, Department of Educational and Counseling Psychology, 16 Hill Hall, University of Missouri, Columbia, MO 65211.).

Heppner, P. P., & Claiborn, C. D. (1989). Social influence research in counseling: A review and critique. *Journal of Counseling Psychology, 36*, 365–387.

Hernstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Herr, E. L., Cramer, S. H., & Niles, S. G. (2004). *Career guidance and counseling through the lifespan: Systematic approaches* (6th ed.). Boston, MA: Allyn & Bacon.

Hess, A. K. (2003). Review of the Myers Briggs Type Indicator(r) Step II (Form Q). In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurement yearbook*. [electronic version]

Hill, C. E. (2014). *Helping skills: Facilitating exploration, insight, and action* (4th ed.). Washington, DC: American Psychological Association.

Hill, C. E., & Lambert, M. J. (2004). Methodological issues in studying psychotherapy processes and outcome. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 84–135). New York: Wiley.

Hiller, J. B., Rosenthal, R., Bornstein, R. F., Berry, D. T. R., & Brunell-Neuleib, S. (1999). A comparative meta-analysis of Rorschach and MMPI validity. *Psychological Assessment, 11*, 278–296.

Hilsenroth, M. J., Ackerman, S. J., Clemence, A. J., Strassle, C. G., & Handler, L. (2002). Effects of structured clinician training on

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

patient and therapist perspectives of alliance early in psychotherapy. *Psychotherapy: Theory/Research/Practice/Training, 39*, 309–323. doi: 10.1037/0033-3204.39.4.309

Hilsenroth, M., Peters, E., & Ackerman, S. (2004). The development of therapeutic alliance during psychological assessment: Patient and therapist perspectives across treatment. *Journal of Personality Assessment, 83*, 332–344.

Hines, M. (1990). Gonadal hormones and human cognitive development. In J. Balthazart (Ed.), *Hormones brains, and behaviors in vertebrates: Sexual differentiation, neuroanatomical aspects, neurotransmitters, and neuropeptides* (pp. 51–63). Basel, Switzerland: Karger.

Hoffman, M. A., Spokane, A. R., & Magoon, T. M. (1981). Effects of feedback mode on counseling outcomes using the Strong-Campbell Interest Inventory: Does the counselor really matter? *Journal of Counseling Psychology*, *28*, 119–125.

Hohenshil, T. H. (1993). Teaching the *DSM-III-R* in counselor education. *Counselor Education and Supervision, 32*, 267–275.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Holland, J. L., Daiger, D., & Power, P. G. (1980). *My Vocational Situation*. Odessa, FL: Psychological Assessment Resources.

Holland, J. L., & Messer, M. A. (2013). *John Holland's SDS Self-Directed Search: Professional manual.* Lutz, FL: PAR.

Honzik, M. P. (1967). Environmental correlates of mental growth: Predictions from the family setting at 21 months. *Child Development, 38*, 323–337.

Hood, A. B., & Johnson, R. W. (2007). *Assessment in counseling: A guide to the use of psychological assessment procedures* (4th ed.). Alexandria, VA: American Counseling Association.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual* (pp. 197–232). New York: Macmillan.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 253–270.

Hunsley, J., & Bailey, J. M. (1999). The clinical utility of the Rorschach: Unfulfilled promises and an uncertain future. *Psychological Assessment, 11*, 266–277.

Hunter, J. E. (1980). *Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB)*. Washington, DC: U.S. Department of Labor.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340–362.

Hunter, J. E., & Schmidt, F. L. (1983). Quantifying the effects of psychological interventions on employee performance and workforce productivity. *American Psychologist, 38*, 473–478.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86*, 721–735.

Jackson, D. N. (1994). *Jackson Personality Inventory—Revised.* Port Huron, MI: Sigma Assessment Systems.

Jackson, D. N. (1996). *Jackson Vocational Interest Survey*. Port Huron, MI: Sigma Assessment Systems.

Jackson, D. N. (1998). *Multidimensional Aptitude Battery-II*. Port Huron, MI: Sigma Assessment System.

Jenkins, J. A. (2009) Strong Interest Inventory assessment tool. In E. A. Whitfield, R. W. Feller, & C. Wood (Eds.), *A counselor's guide to career assessment instruments* (5th ed., pp. 309–319). Broken Arrow, OK: National Career Development Association.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, *39*, 1–23.

Jensen, A. R. (1972). *Genetics and education*. New York: Harper & Row.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: Author. (Mailing address: Joint Committee on Testing Practices, American Psychological Association, 750 First Avenue, NE, Washington, DC, 20002-4242.)

Jones, A. S., & Gelso, C. J. (1988). Differential effects of style of interpretation: Another look. *Journal of Counseling Psychology, 35*, 363–369.

Juhnke, G. A., Juhnke, G. B., & Hsieh, P. (2012). SCATTT: A suicide intervention plan mnemonic for use when clients present suicide intent. *VISTA* (Article 34). Retrieved from http://www.counseling.org/docs/vistas/vistas_2012_article_34.pdf?sfvrsn=3

Kahne, J. (1996). The politics of self-esteem. *American Educational Research Journal, 33*, 3–22.

Kantamneni, N., & Scheel, M. J. (2010). Review of the Strong Interest Inventory [Newly Revised]. In K. F. Geisinger, R. A. Spies, & J. F. Carlson. *The eighteenth mental measurement yearbook*. [electronic version]

Kapes, J. T., & Whitefield, E. A. (2002). *A counselors guide to career assessment instruments* (4th ed.). Tulsa, OK: National Career Development Association.

Kaplan, D. M., Tarvydas, V. M., & Gladding, S. T. (2014). 20/20: A vision for the future of counseling. The new consensus definition of counseling. *Journal of Counseling & Development, 92*, 366372. doi: 10.1002/j.1556-6676.2014.00164.x

Kaplan, R. M., & Saccuzzo, D. P. (2012). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Cengage.

Kaufman, A. S., & Kaufman, N. L. (1983a). *Kaufman Assessment Battery for Children: Administration and scoring manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1983b). *Kaufman Assessment Battery for Children: Interpretive manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1993). *Kaufman Adolescent and Adult Intelligence test manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). San Antonio, TX: Pearson.

Kaufman, A. S., & Kaufman, N. L. (2014). *Kaufman Test of Educational Achievement* (3rd ed.). Bloomington, MN: Pearson.

Kayser, K. (1993). *When love dies: The process of marital disaffection.* New York: Guildford.

Kelley, M. L. (2005). Review of the Piers-Harris Children's Self-Concept Scale. In R. A. Spies, & B. S. Plake (Eds.). *The sixteenth mental measurements yearbook* (2nd ed., pp. 789–790). Lincoln, NE: Buros Institute of Mental Measurements.

Kelley, P. A., Sanspree, M. J., & Davidson, R. C. (2001). Visual impairment in children and youth. In B. Siverstone, M. A. Lang, B. Rosenthal, & E. E. Faye (Eds.), *The lighthouse handbook on visual impairment and visual rehabilitation* (pp. 1137–1151). New York: Oxford University Press.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Kelley, T. L. (1939). The selection of upper and lower group. *Journal of Educational Psychology, 30*, 17–24.

Kelly, K. E. (2010). Review of the Strong Interest Inventory [Newly Revised]. In K. F. Geisinger, R. A. Spies, & J. F. Carlson. *The eighteenth mental measurement yearbook*. [electronic version]

Kent, N., & Davis, D. R. (1957). Discipline in the home and intellectual development. *British Journal of Medical Psychology, 30*, 27–33.

Kirchner, C. (1983). Special education for visually handicapped children: A critique of data on numbers served and costs. *Journal of Visual Impairment & Blindness, 77*, 219–223.

Kiresuk, T. J., Smith, A., & Cardillo, J. E. (1994). *Goal Attainment Scaling: Application, theory, and measurement*. Hillsdale, NJ: Erlbaum.

Knoff, H. M., & Prout, H. T. (1993). *Kinetic drawing system for family and school: A handbook* (5th ed.). Los Angeles: Western Psychological Services.

Knutson, L., & Olson, D. H. (2003). Effectiveness of PREPARE program with premarital couples in community settings. *Marriage and Family, 6*(4), 529546.

Korbin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT® for predicting first-year college grade point average*. New York: The College Board.

Kovacs, M. (2010). *Children's Depression Inventory-2*. North Tonawanda, NY: Multi-Health Systems.

Kristof-Brown, A., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58*(2), 281342. doi: 10.1111/j.1744-6570.2005.00672.x

Krivasty, S. E., & Magoon, T. M. (1976). Differential effects of three vocational counseling treatments. *Journal of Counseling Psychology, 23*, 112–117.

Kuder, Inc. (2012a). *Kuder Career Interests Assessment* (KCIA). Adel, IA: Author.

Kuder, Inc. (2012b). *Kuder Skills Confidence Assessment* (KSCA). Adel, IA: Author.

Kuder, Inc. (2012c). *Kuder Skills Confidence Assessment-Adult* (KSCA-A). Adel, IA: Author.

Kuder, Inc. (2012d) *Kuder Work Values Assessment* (KWVA). Adel, IA: Author.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implication for graduate selection and performance. *Psychological Bulletin, 127*, 162–181. doi: 10.1037/0033-2909.127.1.162

Kuncel, N. R., Ones, D. S. & Sackett, P.R. (2010). Individual differences as predictors of work, education, and broad life outcomes. *Personality and Individual Differences, 49*(4), 331–336. doi: 10.1016/j.paid.2010.03.042

L' Abate, L. (1994). *Family evaluation: A psychological approach*. Thousand Oaks, CA: Sage.

L'Abate, L. (2012). Family assessment: Current and future prospects. In L. Sperry (Ed.), *Assessment of couples and families: Cutting-edge strategies* (pp. 309–328). New York: Brunner-Routledge.

L'Abate, L., Cusinato, M., Maino, E., Colesso,W., & Scilletta, C. (2010). *Relational competence theory: Research and mental health applications*. New York: Springer-Science.

Lachar, D., & Gruber, C. P. (1995). *Personality Inventory for Youth (PIY)*. Los Angeles: Western Psychological Services.

Lachar, D., & Gruber, C. P. (2001). *Personality Inventory for Children* (2nd ed.). Los Angeles: Western Psychological Services.

LaCrosse, M. B. (1980). Perceived counselor social influence and counseling outcome: Validity of the Counselor Rating Form. *Journal of Counseling Psychology, 27*, 320–327.

Lambert, M. J. (2007). Presidential address: What we have learned from a decade of research aimed at improving psychotherapy outcome in routine care. *Psychotherapy Research, 17*, 1–14.

Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice.* Washington, DC: American Psychological Association.

Lambert, M. J., Bailey, R. J., Kimball, K., Shimokawa, K., Harmon, S. C., & Slade, K. (2008). *CSTs manual-Brief version–40.* Unpublished manuscript, Department of Psychology, Brigham Young University, Provo, UT.

Lambert, M. J., Bergin, A. E., & Collins, J. L. (1977). Therapist induced deterioration in psychotherapy patients. In A. S. Gurman & A. M. Razin (Eds.), *Effective psychotherapy: A handbook of research* (pp. 452–481). New York: Pergamon Press.

Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69*, 159–172.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., & Burlingame, G. M. (2004). *Administration and scoring manual for the Outcome Questionnaire-45*. Salt Lake City, UT: OQ Measures.

Lambert, M. J., & Vermeersch, D. A. (2013). Psychological assessment in treatment. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology, Volume 2* (pp. 213229). Washington, DC: American Psychological Association.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*, 288–301.

Lanning, K. (2003). Review of the Myers Briggs Type Indicator(r) Step II (Form Q). In B. S. Plake, J. C. Impara, & R. A. Spies (Eds.), *The fifteenth mental measurement yearbook*. [electronic version]

Lapan, R. T., & Kosciulek, J. F. (2001). Toward a community career system program evaluation framework. *Journal of Counseling & Development, 79*, 3–16.

Larson, J. H. (2002). *Consumer update: Marriage preparation*. [Brochure]. Alexandria, VA: American Association of Marriage and Family Therapists.

Larson, J. H., Vatter, R. S., Galbraith, R. C., Holman, T. B., & Stahmann, R. F. (2007). The Relationship Evaluation (RELATE) with therapist-assisted interpretation: Short term effects on premarital relationships. *Journal of Marital and Family Therapy, 28*, 233–239.

Lazarus, S. S., Thurlow, M. L., Lail, K. E., & Christensen, L. (2009). Longitudinal analysis of state accommodations polices: Twelve years of change, 1993–2005. *Journal of Special Education, 43*, 67–80. doi: 10.1177/0022466907313524

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports, 68*, 203–210.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*, 79–122.

Leong, F. T. L., & Hartung, P. J. (2000). Cross-cultural career assessment: Review and prospects for the new millennium. *Journal of Career Assessment, 8*, 391–401.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Lerner, B. (1981). The minimum competence testing movement: Social, scientific, and legal implications. *American Psychologist, 36*, 1056–1066.

Lesher, E. L., & Berryhill, J. S. (1994). Validation of the Geriatric Depression Scale—Short Form among inpatients. *Journal of Clinical Psychology, 50*, 256–260.

Lichtenberger, E. O., Mather, N., Kaufman, N. L., & Kaufman, A. S. (2004). *Essentials of assessment report writing*. Hoboken, NJ: Wiley.

Lichtenberg, J. W., & Hummel, T. J. (1998). The communication of probabilistic information through test interpretation. In C. Claiborn (Chair), *Test interpretation in counseling—Recent research and practice*. Symposium conducted at the conference of the American Psychological Association, San Francisco.

Lindzey, G. (1959). On the classification of projective techniques. *Psychological Bulletin, 56*, 158–168.

Linn, R. L. (1994). Fair test use: Research and policy. In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 363–375). Hillsdale, NJ: Erlbaum.

Liu, W. M., & Watt, S. K. (2013). Counseling and the culture of economic disadvantage. In C. C. Lee (Ed.), *Multicultural issues in counseling: New approaches to diversity* (4th ed., pp. 259–274). Alexandria, VA: American Counseling Association.

Locke, H., & Wallace, K. (1959). Short marital adjustment and predictions tests: Their reliability and validity. *Marriage and Family Living, 2*, 251–255.

Loehlin, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist, 10*, 1285–1292.

Lovett, B. J., & Lewandowski, L. J. (2015). *Testing accommodations for students with disabilities: Research-based practice.* Washington, DC: American Psychological Association.

Lopez, M. N., Charter, R. A., Mostafavi, B., Nibut, L. P., & Smith, W. E. (2005). Psychometric prosperities of the Folstein Mini-Mental State Examination. *Assessment, 12*, 137–144. doi: 10.1177/1073191105275412

Lopez, S. J., Snyder, C. R., & Rasmussen, H. N. (2003). Striking a vital balance: Developing a complementary focus on human weakness and strength through positive psychological assessment. In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures* (pp. 3–20). Washington, DC: American Psychological Association.

Luria, A. R. (1966). *Human brain and psychological processes*. New York: Harper & Row.

Lutz, W., Lambert, M. J., Harmon, S. C., Stulz, N., Tschitsaz, A., & Schurch, E. (2006). The probability of treatment success, failure, and duration.What can be learned from empirical data to support decision making in clinical practice? *Clinical Psychology and Psychotherapy, 13*, 223–232. doi: 10.1002/cpp.496

MacCluskie, K. C., Welfel, E. R., & Toman, S. M. (2002). *Using test data in clinical practice: A handbook of mental health professionals*. Thousand Oaks, CA: Sage.

Machover, K. (1949). *Personality projection in the drawings of the human figure: A method of personality investigation*. Springfield, IL: Charles C. Thomas.

Manges, K. J. (2001). Review of the Family Assessment Measure, Version III. In B. Plake & J. Impara (Eds.), *The fourteenth mental measurements yearbook* (pp. 480–482). Lincoln, NE: Buros Institute of Mental Measures.

Manuele-Adkins, C. (1989). Review of the Self-Directed Search. In J. C. Conoley & J. J. Kramer (Eds.), *Tenth mental measurements yearbook* (pp. 738–740). Lincoln, NE: The University of Nebraska Press.

Maraist, C. C., & Russell, M. T. (2002). *16PF Fifth Edition norm supplement, release 2002*. Champaign, IL: Institute for Personality and Ability Testing.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics, 23*, 337–341.

Mastrangelo, P. M. (2001). Review of the Myers-Briggs Type Indicator. In B. Plake & J. Impara (Eds.), *The fourteenth mental measurements yearbook* (pp. 714–715). Lincoln, NE: Buros Institute of Mental Measures.

Maruish, M. E. (2004). Introduction. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Volume 1 general considerations* (3rd ed., pp. 1–64). Mahwah, NJ: Erlbaum.

Mayfield, D. G., McLeod, G., & Hall, P. (1974). The CAGE questionnaire: Validation of a new alcoholism screening instrument. *American Journal of Psychiatry*, *131*, 1121–1123.

McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist, 52*, 509–516.

McCrae, R. R., & Costa, P. T. (2010). *NEO inventories: Professional manual*. Lutz, FL: PAR.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175–215.

McGhee, R. L., Ehrler, D. J., & Buckhalt, J. A. (2007). *FFPI-C: Five Factors Personality Inventory-Children examiner's manual*. Austin, TX: PRO-ED.

McGoldrick, M., Gerson, R., & Petry, S. (2008). *Genograms: Assessment and intervention* (3rd ed.). New York: W. W. Norton.

McGrew, K. S. (1997). Analysis of major intelligence batteries according to a proposed comprehensive Gc-Gf framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–180). New York: Guilford.

McGue, M., Bouchard, T. J., Jr., Iacono, W. G., & Lykken, D. T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin & G. E. McClearn (Eds.), *Nature, nurture, & psychology* (pp. 59–76). Washington, DC: American Psychological Association.

McMahon, M., Patton, W., & Watson, M. (2003). Developing qualitative career assessment processes. *Career Development Quarterly, 51*, 194–202.

McMichael, A. J., Baghurst, P. A., Wigg, N. R., Vimpani, G. V., Robertson, E. F., & Roberts, R. J. (1988). Port Pirie cohort study: Environmental exposure to lead and children's abilities at the age of four years. *New England Journal of Medicine, 319*, 468–475.

McNamara, K. (1992). Depression assessment and intervention: Current status and future direction. In S. D. Brown & R. W. Lent (Eds.), *Handbook of counseling psychology* (2nd ed., pp. 691–718). New York: Wiley.

McShane, D. A., & Plas, J. M. (1984). The cognitive functioning of American Indian children: Moving from the WISC to the WISC-R. *School Psychology Review, 13*, 61–73.

Meijer, R. R., de Vries, R. M., & van Bruggen, V. (2011). An evaluation of the Brief Symptom Inventory-18 using item response theory: Which items are most strongly related to psychological distress? *Psychological Assessment, 23*, 193–202. doi: 10.1037/a0021292

Meir, S. T. (2003). *Bridging case conceptualization, assessment, and intervention*. Thousand Oaks, CA: Sage.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Merrell, K. W. (2007). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). Mahwah, NJ: Erlbaum.

Merriam-Webster (2014). *Merriam-Webster dictionary*. Retrieved http://www.merriam-webster.com/

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.

Michel, D. M. (2002). Psychological assessment as a therapeutic intervention in patients hospitalized with eating disorders. *Professional Psychology: Research and Practice, 33*, 470–477.

Miller, G. A. (1997). *The Substance Abuse Subtle Screening Inventory 3 (SASSI-3)*. Springville, IN: The SASSI Institute.

Miller, G. A. (2001). *The Adolescent Substance Abuse Subtle Screening Inventory-A2 (SASSI-A2)*. Springville, IN: The SASSI Institute.

Miller, M. D. (2010). Review of the Wechsler Individual Achievement Test- Third Edition. In K. F. Geisinger, R. A. Spies, & J. F. Carlson. *The eighteenth mental measurement yearbook*. [electronic version]

Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The Partners for Change Outcome System. *Journal of Clinical Psychology, 61*, 199–208.

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2nd ed.). New York: Guilford Press.

Miller, W. R., & Rose, G. S. (2009). Toward a theory of motivational interviewing. *American Psychologist, 64*, 527–537.

Millon, T. (1969). *Modern psychological pathology: A biosocial approach to maladaptive learning and functioning*. Philadelphia: Saunders.

Millon, T. (1994). *Millon Index of Personality Styles (MIPS) manual*. San Antonio, TX: Psychological Corporation.

Millon, T. (2006). *Millon Clinical Multiaxial Inventory III manual* (3rd ed.). Minneapolis: Pearson Assessments.

Millon, T., Davis, R., & Millon, C. (1997). *Millon Clinical Multiaxial Inventory III manual* (2nd ed.). Minneapolis: National Computer Systems.

Millon, T., Millon, C., & Davis, R. (1994). *Millon Adolescent Clinical Inventory*. Minneapolis: National Computer Systems.

Miner, C. U., & Sellers, S. T. (2002). Career Assessment Inventory. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (4th ed., pp. 204–209). Tulsa, OK: National Career Development Association.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 362. doi: 10.1207/S15366359MEA0101_0

Mitchell, R. R., & Friedman, H. (1994). *Sandplay: Past, present, and future*. New York: Routledge.

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology: Science and Practice, 2*, 1–27.

Moos, R. H. (1973). Conceptualizing human environments. *American Psychologist, 28*, 652–665.

Moos, R. H., & Moos, B. S. (1994). *Family Environment Scale manual: Development, application, research*. Palo Alto, CA: CPP, Inc.

Morris, T. M. (1990). Culturally sensitive family assessment: An evaluation of the Family Assessment Device with Hawaiian-American and Japanese-American families. *Family Process, 29*, 105–116.

Morrison, J. (2007). *Diagnosis made easier: Principles and techniques for mental health clinicians*. New York: Guilford.

Morrison, J. (2014). *DSM-5 made easy: The clinician's guide to diagnosis*. New York: Guilford.

Murphy, E., & Meisgeier, C. (1987). *Murphy-Meisgeier Type Indicatory for Children*. Palo Alto, CA: CPP, Inc.

Murphy, L. L., Geisinger, K F., Carlson, J. F., & Spies, R. A. (2011). *Tests in print VIII*. Lincoln, NB: Buros Center for Testing.

Murray, H. A. (1943). *Thematic Apperception Test*. Cambridge, MA: Harvard University Press.

Murrie, D., Martindale, D. A., & Epstein, M. (2009). Unsupported assessment techniques in child sexual abuse evaluations. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony* (pp. 397–420). Hoboken, NJ: Wiley.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Palo Alto, CA: CPP, Inc.

Myers, I. B., with Myers, P. B. (1980). *Gifts differing: Understanding personality types*. Palo Alto, CA: CPP, Inc.

Naglieri, J. A., & Das, J. P. (1997). *Das-Niglieri: Cognitive Assessment System: Interpretive handbook*. Itasca, IL: Riverside.

Naglieri, J. A., & Das, J. P. (2005). Planning, Attention, Simultaneous, and Successive (PASS) theory: A revision of the concept of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 120–135). New York: Guilford.

National Board for Certified Counselors. (2014). *Content covered in the NCE.* Retrieved from http://nbcc.org/InnerPageLinks/ContentCoveredInTheNCE.

National Career Development Association. (1997). *NCDA guidelines for the use of the Internet for the provision of career information and planning services*. Columbus, OH: Author.

National Career Development Association. (2007). *National Career Development Association code of ethics*. Broken Arrow, OK: Author.

National Center on Response to Intervention. (2010). *Essential components of RTI—A closer look at response to intervention*. Retrieved from http://www. rti4success.org/pdf/rtiessentialcomponents_042710.pdf

National Research Council. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–98.

Nelson, R. O. (1983). Behavioral assessment: Past, present, and future. *Behavioral Assessment, 5*, 196–206.

Nevill, D. D., & Super, D. E. (1989). *The Values Scale: Theory, application, and research*. Palo Alto, CA: CPP, Inc.

Newman, M. L., & Greenway, P. (1997) Therapeutic effects of providing MMPI-2 test feedback to clients at a university counseling service: A collaborative approach. *Psychological Assessment, 9*, 122–131.

Newmark, C. S., & McCord, D. M. (1996). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2). In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 1–58). Boston: Allyn & Bacon.

Nguyen, T. D., Attkisson, C. C., & Stegner, B. L. (1983). Assessment of patient satisfaction: Development and refinement of a Service Evaluation Questionnaire. *Evaluation and Program Planning, 6*, 299–313.

Nichols, D. S. (1992). Review of the Minnesota Multiphasic Personality Inventory 2. In J. C. Conoley & J. J. Kramer (Eds.), *Eleventh mental measurements yearbook* (pp. 562–565). Lincoln, NE: The University of Nebraska Press.

Nicholson, C. L., & Hibpshman, T. H. (2002). *Slosson Intelligence Test for Children and Adults,* (rev. ed.). San Antonio, TX: Pearson.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Nurse, A. R. (1999). *Family assessment: Effective uses of personality tests with couples and families*. New York: Wiley.

Nurse, A. R., & Sperry, L. Standardized assessments: In L. Sperry (Ed.), *Family assessment: Contemporary and cutting-edge strategies* (2nd ed., pp.5381). New York: Routledge.

Oetting, E. R., & Beauvais, F. (1990). Adolescent drug use: Findings of national and local surveys. *Journal of Consulting and Clinical Psychology, 58*, 385–394.

Ogles, B. M. (2013). Measuring change in psychotherapy research. In M. J. Lambert (Ed.), *Bergin & Garfield's Handbook of psychotherapy and behavior change* (6th ed.). New York: Wiley.

Ogles, B. M., Lambert, M. J., & Fields, S. A. (2002). *Essentials of outcome assessment*. New York: Wiley.

Okun, B. F., & Kantrowitz, R. E. (2007). *Effective helping: Interviewing and counseling techniques* (7th ed.). Belmont, CA: Brooks/Cole.

Olin, D. W., & Blow, F. C. (2000). Substance use dis-orders in late life. In I. Katz & D. Oslin (Eds.), *Annual review of gerontology and geriatrics: Focus on psychopharmacologic interventions in late life* (pp. 213–234). New York: Springer.

Oliver, L. W., & Spokane, A. R. (1988). Career-intervention outcome: What contributes to client gain? *Journal of Counseling Psychology, 35*(4), 447462. doi: 10.1037/0022-0167.35.4.447

Oliver, L. W., & Zack, J. S. (1999). Career assessment on the Internet: An exploratory study. *Journal of Career Assessment*, *7*, 323–356.

Olson, D. H., DeFrain, J. & Skogrand, L. (2010). *Marriages and families: Intimacy, diversity, and strengths* (7th ed.). McGraw-Hill.

Olson, D. H., & Gorall, D. M. (2006). *FACES IV & the Circumplex Model*. Retrieved from http://www. facesiv.com/pdf/3.innovations.pdf

Olson, D. H., Gorall, D. M., & Tiesel, J. (2004). *Family Adaptability and Cohesion Evaluation Scale IV*. Roseville, MN: Life Innovations.

Olson, D. H., Portner, J., & Lavee, Y. (1985). *FACES III: Family Adaptability and Cohesion Scale*. St. Paul, MN: University of Minnesota, Department of Family Social Science.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2004). Cognitive ability in selection decisions. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 431–468). Thousand Oaks, CA: Sage.

Oosterhof, A. (2000). *Classroom application of educational measurement* (3rd ed.). Upper Saddle River, NJ: Merrill Prentice-Hall.

Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 307–389). New York: Wiley.

Osborne, W. L., Brown, S., Niles, S., & Miner, C. U. (1997). *Career development, assessment, and counseling: Applications of the Donald E. Super C-DAC approach*. Alexandria, VA: American Counseling Association.

Osipow, S. H. (1987). *Career Decision Scale manual*. Odessa, FL: Psychological Assessment Resources.

Osipow, S. H., & Gati, I. (1998). Construct and concurrent validity of the Career Decision-Making Difficulties Questionnaire. *Journal of Career Assessment*, *6*, 347–364.

Oster, G. D., & Crone, P. G. (2004). *Using drawings in assessment and therapy: A guide for mental health professionals* (2nd ed.). New York: Brunner-Routledge.

Oswald, D. P. (2005). Review of the Piers-Harris Children's Self-Concept Scale, (Second Edition.) In R. A. Spies, & B. S. Plake (Eds.). *The sixteenth mental measurements yearbook* (2nd ed., pp. 790–792). Lincoln, NE: Buros Institute of Mental Measurements.

Padilla, A. M., & Borsato, G. N. (2008). Issues in culturally appropriate psychoeducational assessment. In L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (3rd ed., pp. 5–21). Hoboken, NJ: Wiley.

Parsons, F. (1909). *Choosing a vocation*. Boston, MA: Houghton Mifflin.

Patrick, J., Blosel, C. W., & Gross, C. L. (2008). Armed Services Vocational Aptitude Battery/Career exploration. In E. A Whitfield, R. W. Feller, & C. Wood (Eds.), *A counselor's guide to career assessment instruments* (5th ed., pp. 97–104). Broken Arrow, OK: National Career Development Association.

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980) Validity generalization results for tests used to predict training success and job proficiency. *Journal of Applied Psychology, 65*, 373–406.

Percevic, R., Lambert, M. J., & Kordy, H. (2006). What is the predictive value of responses to psychotherapy for its future course? Empirical explorations and consequences for outcome monitoring. *Psychotherapy Research, 16*, 364373. doi: 10.1080/10503300500485524

Peterson, C. H., Lomas, G. I., Neukrug, E. S., & Bonner, M. W. (2014). Assessment use by counselors in the United States: Implications for policy and practice. *Journal of Counseling & Development, 92*(1), 9098. doi: 10.1002/j.1556-6676.2014.00134.x

Phillips, S. E. (2000). *GI Forum v. Texas Education Agency*: Psychometric evidence. *Applied Measurement in Education*, *13*, 343–385.

Piaget, J. (1972). *The psychology of intelligence*. Totowa, NJ: Littlefield Adams.

Piers, E. V., Harris, D. B., & Herzberg, D. S. (2002). *Piers-Harris Children's Self-Concept Scale* (2nd ed.). Los Angeles: Western Psychological Services.

Piotrowski, C. (1999). Assessment practices in the era of managed care: Current status and future directions. *Journal of Clinical Psychology, 55*, 787–796.

Polanski, P. J., & Hinkle, J. S. (2000). The mental status examination: Its use by professional counselors. *Journal of Counseling & Development, 78*, 357–364.

Ponterotto, J. G., Gretchen, D., & Chauhan, R. V. (2001). Cultural identity and multicultural assessment: Quantitative and qualitative. In L. A. Suzuk, J. G. Ponterrotto, & P. J. Meller (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (2nd ed., pp. 67–99). San Francisco: Jossey-Bass.

Ponterotto, J. G., Pace, J. A., & Kavan, K. (1989). A counselor's guide to the assessment of depression. *Journal of Counseling and Development, 67*, 301–309.

Pope, M. (2009). O*NET Interest Profiler and Computerized O*NET Interest Profiler (CIP). In E. A. Whitfield, R. W. Feller, & C. Wood (Eds.), *A counselor's guide to career assessment instruments* (5th ed., pp. 274–285). Broken Arrow, OK: National Career Development Association.

Poston, J. M., & Hanson, W. E. (2010). Meta-analysis of psychological assessment as a therapeutic intervention. *Psychological Assessment, 22*, 203–212.

Powers, D. E., & Rock, D. A. (1999). Effects of coaching on SATI: Reasoning test scores. *Journal of Educational Measurement, 36*, 93–118.

Poznanski, E. O., & Mokros, H. B. (1996). *Children's Depression Rating Scale (*rev. ed.*)*. Los Angeles: Western Psychological Services.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Pratt, S. L., & Moreland, K. L. (1998). Individuals with other characteristics. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. D. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 349–371). Washington, DC: American Psychological Association.

Prediger, D. J. (1999). Integrating interests and abilities for career exploration: General considerations. In M. L. Savickas & A. R. Spokane (Eds.), *Vocational interests: Meaning, measurement, and counseling use* (pp. 295–325). Palo-Alto, CA: Davies-Black.

Prince, J. P., & Heiser, L. J. (2000). *Essentials of career interest assessment*. New York: John Wiley.

Pryor, R. G. L., & Taylor, N. B. (1986). On combining scores from interest and values measures in counseling. *Vocational Guidance Quarterly, 34*, 178–187.

Psychological Corporation. (1991). *Differential Aptitude Tests (*5th ed.*) Career Interest Inventory counselors' manual*. San Antonio, TX: Author.

Pugh, R. C. (1998). Review of the Campbell Interest and Skill Inventory. In C. W. Conoley & J. Impara (Eds.), *The thirteenth mental measurements year book* (pp. 167–170). Lincoln, NE: Buros Institute of Mental Measures.

Rada, R. (2003). *HIPAA@IT reference, 2003 edition: Health information transactions, privacy, and security*. [Electronic resource]. Norwood, MA: Hypermedia Solutions Limited.

Raven, J. C., Raven, J., Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Standard progressive matrices*. Bloomington, MN: Pearson.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children* (2nd ed.). Bloomington, MN: Pearson.

Reynolds, W. M. (1988). *Suicidal Ideation Questionnaire*. Odessa, FL: Psychological Assessment Resources.

Reynolds, W. M. (1989). *Reynolds Child Depression Scale*. Lutz, FL: Psychological Assessment Resources.

Reynolds, W. M. (1991). *Adult Suicidal Ideation Questionnaire*. Lutz, FL: Psychological Assessment Resources.

Reynolds, W. M. (2002). *Reynolds Depression Scale* (2nd ed.). Lutz, FL: Psychological Assessment Resources.

Reynolds, W. M., & Kobak, K. A. (1995). *Hamilton Depression Inventory*. Odessa, FL: Psychological Assessment Resources.

Rivera, L. M. (2008). Acculturation and multicultural assessment: Issues, trends, and practice. In L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (3rd ed., pp. 73–91). Hoboken, NJ: Wiley.

Roberts, B. W., & Robins, R. W. (2000). Broad dispositions, broad aspirations: The intersection of personality traits and major life goals. *Personality and Social Psychology Bulletin, 26*, 1284–1296.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*(1), 1–29. doi:10.10370033-2909.132.1.1

Robertson, G. J. (2002). *Wide Range Achievement Test-Expanded*. Lutz, FL: Psychological Assessment Resources.

Rogers, J. R., Guelulette, C. M., Abbey-Hines, J., Carney, J. V., & Werth, J. L., Jr. (2001). Rational suicide: An empirical investigation of counselor attitudes. *Journal of Counseling & Development, 79*, 365–372.

Rogers, R., Sewell, K. W., Harrison, K. S., & Jordan, M. J. (2006). The MMPI-2 restructured clinical scales: A paradigmatic shift in scale development. *Journal of Personality Assessment, 87*, 139–147.

Rohrbraugh, M., Rogers, J. C., & McGoldrick, M. (1992). How do experts read family genograms? *Family Systems Medicine, 10*, 79–89.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scale* (5th ed.). Itasca, IL: Riverside Publishing.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Thousand Oaks, CA: Sage.

Roszkowski, M. J. (2001). Review of the Jackson Vocational Interest Survey [1995 Revision]. In B. Plake & J. Impara (Eds.), *The fourteenth mental measurements yearbook* (pp. 615–618). Lincoln, NE: Buros Institute of Mental Measures.

Roth, L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297–330.

Rotter, J. B., Lah, M. I., & Rafferty, J. E. (1992). *Rotter Incomplete Sentence Blank* (2nd ed.). San Antonio, TX: The Psychology Corporation.

Rounds, J., & Tracey, T. J. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology, 43*(3), 310329. doi: 10.1037/0022-0167.43.3.310

Rounds, J. B. (1990). The comparative and combined utility of work value and interest data in career counseling with adults. *Journal of Vocational Behavior, 37*, 32–45.

Rounds, J. B., & Armstrong, P. I. (2005). Assessment of needs and values. In S. D. Brown & R. W. Lent, (Eds.), *Career development and counseling* (pp. 305–329). Hoboken, NJ: Wiley.

Ryan Krane, N. E., & Tirre, W. C. (2005). Ability assessment in career counseling. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 330–352). Hoboken, NJ: Wiley.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of scoring adjustment in preemployment testing. *American Psychologist, 49*, 929–954.

Sampson, J. P. Jr. (2009). Modern and postmodern career theories: An unnecessary divorce. *Career Development Quarterly, 58*, 91–96.

Sampson, J. P. Jr., Carr, D. L., Lumsden, J. A., Smission, C., & Dozier, C. (2009). Computer-assisted career assessment: State of the art. In E. A. Whitfield, R. W. Feller, & C. Wood (Eds.), *A counselor's guide to career assessment instruments* (5th ed., pp. 43–60). Broken Arrow, OK: National Career Development Association.

Sampson, J. P. Jr., Lumsden, J. A., & Carr, D. L. (2002). Computer-assisted career assessment. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (4th ed., pp. 47–63). Tulsa, OK: National Career Development Association.

Sampson, J. P. Jr., & Osborne, D. S. (2015). Using information and communication technology in delivering career interventions. In P. J. Hartung, M. L. Savickas, & W. B. Walsh (Eds.), *APA handbook of career interventions. Vol. 2 applications* (pp. 5770). Washington, DC: American Psychological Association.

Sampson, J. P. Jr., Peterson, G. W., Lenz, J. G., Reardon, R. C., & Saunders, D. E. (1996). *Career Thoughts Inventory*. Lutz, FL: PAR Inc.

Sanchez, H. B. (2001). Risk factor model for suicide assessment and intervention. *Professional Psychology: Research and Practice, 32*, 351–358.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations (*5th ed.). San Diego, CA: Jerome M. Sattler Publisher.

Sattler, J. M., & Evans, C. A. (2014). Visual impairments. In J. M. Sattler (Ed.), *Foundations of behavioral, social, and clinical assessment of children* (6th ed., pp.557–575). San Diego, CA: Jerome M. Sattler, Publisher.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Sattler, J. M., & Garro, A. (2014). Interviewing children, parents, teachers, and families. In J. M. Sattler (Ed.), *Foundations of behavioral, social, and clinical assessment of children* (6th ed., pp.205–234). San Diego, CA: Jerome M. Sattler, Publisher.

Sattler, J. M., & Riddell, R. P. (2014). Observational methods, part I. In J. M. Sattler (Ed.), *Foundations of behavioral, social, and clinical assessment of children* (6th ed., pp. 257–293). San Diego, CA: Jerome M. Sattler, Publisher.

Savickas, M. L. (2001). Toward a comprehensive theory of career development: Dispositions, concerns, and narratives. In F. L. T. Leong & A. Barak (Eds.), *Contemporary models in vocational psychology: A volume in honor of Samuel H. Osipow* (pp. 295–320). Mahwah, NJ: Erlbaum.

Savickas, M. L. (2005). The theory and practice of career construction. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling* (pp. 42–70). Hoboken, NJ: Wiley.

Savickas, M. L. (2011a). Career construction theory and practice. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (2nd ed., pp. 147–186). Hoboken, NJ: Wiley.

Savickas, M. L. (2011b). *Career counseling.* Washington, DC: American Psychological Association.

Savickas, M. L., & Porfeli, E. J. (2011). Revision of the career maturity inventory: The adaptability form. *Journal of Career Assessment, 19*(4), 355–374. doi: 10.1177/106907271140934

Savickas, M. L., & Porfeli, E. J. (2012). Career adapt-abilities scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior, 80*(3), 661–673. doi: 10.1016/j.jvb.2012.01.011

Savickas, M. L., Taber, B. J., & Spokane, A. R. (2002). Convergent and discriminant validity of five interest inventories. *Journal of Vocational Behavior, 61*, 139–184.

Scarr, S. (1992). Developmental theories for the 1990s: Developmental and individual differences. *Child Development, 63*, 1–19.

Scarr, S. (1993). Biological and cultural diversity: The legacy of Darwin for development. *Child Development, 64*, 1333–1354.

Schaie, K. W. (1996). Intellectual development in adulthood. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging (*4th ed., pp. 266–286). San Diego, CA: Academic Press.

Schenck, P. M. (2009). Kuder Career Planning System. In E. A. Whitfield, R. W. Feller, & C. Wood (Eds.), *A counselor's guide to career assessment instruments* (5th ed., pp. 163–173). Broken Arrow, OK: National Career Development Association.

Schrank, F.A., McGrew, K.S., Mather, N., & Woodcock, R.W. (2014). *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside Publishing.

Schuerger, J. M. (2001). *16PF Adolescent Personality Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.

Scott, C. L. (2010). The Americans with Disabilities Act Amendments Act of 2008: Implications for the forensic psychiatrist. *Journal of the American Academy of Psychiatry and the Law, 38*, 95–99.

Salend, S. J. (2009). *Classroom testing and assessment for all students: Beyond standardization*. Thousand Oaks, CA: Corwin.

Schaefer, C. E., Gitlin, K., & Sandgrund, A. (Eds.). (1991). *Play diagnosis and assessment*. New York: Wiley.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540.

Schmidt, F. L., Hunter, J. E., & Caplan, J. R. (1981). Validity generalization results for two jobs in the petroleum business. *Journal of Applied Psychology, 66*, 261–273.

Schoenrade, P. (2002). Values Scale. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (4th ed., pp. 299–302). Tulsa, OK: National Career Development Association.

Schwab, D. P., & Packard, G. L. (1973). Response set distortion on the Gordon Personal Inventory and the Gordon Personal Profile in the selection context: Some implications for predicting employee behavior. *Journal of Applied Psychology, 58*, 372–374.

Schwiebert, V. L. (2009). Selecting a career assessment instrument. In E. A. Whitfield, R. W. Feller, & C. Wood (Eds.), *A counselor's guide to career assessment instruments* (5th ed., pp. 127–133). Broken Arrow, OK: National Career Development Association.

Sireci, S. G., & Geisinger, K. F. (1998). Equity issues in employment testing. In J. Sandoval, C. L. Frisby, K. F. Geisinger, J. D. Scheuneman, & J. R. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 105–140). Washington, DC: American Psychological Association.

Skiba, R. J., Poloni-Staudinger, L., Gallini, S., Simmons, A. D., & Feggins-Azziz, R. (2006). The disproportionality of African American students with disabilities across educational environments. *Exceptional Children, 72*, 411–424.

Skinner, H. A., Steinhauer, P. D., & Santa-Barbara, J. (1995). *The Family Assessment Measure*. North Tonawanda, NY: Multi-Health Systems.

Slaney, R. B., & MacKinnon-Slaney, F. (1990). The use of vocational card sorts in career counseling. In M. L. Savickas & W. B. Walsh (Eds.), *Handbook of career counseling theory and practice* (pp. 317–371). Palo Alto, CA: Davies-Black.

Slaney, R. B., & Suddarth, B. H. (1994). The Values Scale (VS). In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (pp. 236–240). Alexandria, VA: National Career Development Association.

Snyder, C. R. (1997). *Marital Satisfaction Inventory, Revised manual*. Los Angeles: Western Psychological Services.

Snyder, D. K., Heyman, R. E., & Hayes, S. N. (2009). Assessing couples. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 457–484). New York: Oxford University Press.

Sodano, S. M. (2015). Meaning, measurement, and assessment of vocational interests for career intervention. In P. J. Hartung, M. L. Savickas, & W. B. Walsh (Eds.), *APA handbook of career interventions* (pp. 281–301). Washington, DC: American Psychological Association.

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage or similar dyads. *Journal of Marriage and the Family, 38*, 15–28.

Spearman, C. (1927). *The abilities of man*. New York: Macmillan.

Spengler, P. M., Strohmer, D. C., Dixon, D. N., & Shivy, V. A. (1995). A scientist-practitioner model of psychological assessment: Implications for training, practice, research. *The Counseling Psychologist, 23*, 506–534.

Spielmans, G. I., Masters, K. S., & Lambert, M. J. (2006). A comparison of rational versus empirical methods in prediction of negative psychotherapy outcome. *Clinical Psychology and Psychotherapy, 13*, 202–214. doi: 10.1002/cpp.491

Spillane, S. A. (2001). Review of the Family Assessment Measure Version III. In B. Plake & J. Impara (Eds.), *The fourteenth mental measurements yearbook* (p. 482). Lincoln, NE: Buros Institute of Mental Measures.

Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence,

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

personality and motivation? *Educational Research, 56*(2), 230–243. doi: 10.1080/00131881.2014.898917

Spokane, A. R. (1991). *Career interventions*. Englewood Cliffs, NJ: Prentice Hall.

Spokane, A. R., & Jacob, E. J. (1996). Career and vocational assessment 1993–1994: A biennial review. *Journal of Career Assessment, 4*, 1–32.

Steer, R. A., & Clark, D. A. (1997). Psychometric characteristics of the Beck Depression Inventory II. *Measurement and Evaluation in Counseling and Development*, *30*, 128–136.

Stein. S. (2007). Review of the Behavior Assessment System for Children, Second Edition. In R. A. Spies, B. S. Plake, K. F. Geisinger, & J. F. Carlson. *The seventeenth mental measurement yearbook*. [electronic version]

Stelmachers, Z. T. (1995). Assessing suicidal clients. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 367–379). New York: Oxford University Press.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence*. New York: Viking.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conception of intelligence. *Journal of Personality and Social Psychology, 41*, 37–55.

Sternberg, R. J., & Grigorenko, E. L. (2008). Ability testing across cultures. In L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (3rd ed., pp. 449–470). Hoboken, NJ: Wiley.

Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education, 13*, 275–287.

Stratton, K., Howey, C., & Battaglia, F. (1996). *Fetal alcohol syndrome: Diagnosis, epidemiology, prevention, and treatment*. Washington, DC: National Academy Press.

Strickler, L. J. (1969). "Test-wiseness" on personality scales. *Journal of Applied Psychology Monograph, 53*(3) 1–17.

Su, R., Rounds, J., & Armstrong, P. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*(6), 859–884. doi:10.1037/a0017364

Substance Abuse and Mental Health Services Administration. (2010). *Results from the 2009 National Survey on Drug Use and Health: Volume I. Summary of National Findings* (Office of Applied Studies, NSDUH Series H-38A, HHS Publication No. SMA 10-4586 Findings). Rockville, MD: Office of Applied Studies.

Super, D. E., Thompson, A. S., Lindeman, R. H., Jordaan, J. P., & Myers, R. A. (1988a). *Adults Career Concerns Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Super, D. E., Thompson, A. S., Lindeman, R. H., Jordaan, J. P., & Myers, R. A. (1988b). *Career Development Inventory*. Palo Alto, CA: CPP, Inc.

Suzuki, L. A., Alexander, C. M., Lin, P., & Duffy, K. M. (2006). Psychopathology in the schools: Multicultural factors that impact assessment and intervention. *Psychology in the Schools, 43*, 329-438. doi: 10.1002/pits.20157

Suzuki, L. A., Short, E. L., Pieterse, A., & Kugler, J. (2001). Multicultural issues and the assessment of aptitude. In L. A. Suzuki, J. G. Ponterotto, & P. J. Mellers (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (2nd ed., pp. 359–382). San Francisco: Jossey-Bass.

Swanson, J. L. (2012). Work and psychological health. In N. A. Fouad & J. A. Carter (Eds.). *APA handbook of counseling psychology* (pp. 3–28). Washington, DC: American Psychological Association.

Swenson, C. H. (1968). Empirical evaluations of human figure drawings: 1957–1966. *Psychological Bulletin, 70*, 20–44.

Tellegen, A., Ben-Porath, Y. S. (2008). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Technical manual*. Minneapolis: University of Minnesota Press.

Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *The MMPI-2 restructured clinical (RC) scales: Development, validation, and interpretation*. Minneapolis: University of Minnesota Press.

Tharinger, D. J., Finn, S. E., Hersh, B., Wilkinson, A., Christopher, G. B., & Tran, A. (2008). Assessment feedback with parents and preadolescent children: A collaborative approach. *Professional Psychology: Research and Practice, 39*, 600–609. doi: 10.1037/0735-7028.39.6.600

Thomlison, B. (2010). *Family assessment handbook: An introduction and practical guide to family assessment* (3rd ed.). Belmont, CA: Brooks/Cole.

Thompson, B., & Snyder, P. A. (1998). Statistical significance and reliability analyses in recent. *Journal of Counseling and Development, 76*, 436–441.

Thorndike, R. M. (2005). Review of the Kaufman Assessment Battery for Children. In Spies, R. A., & Plake, B. S. (Eds.). *The sixteenth mental measurements yearbook* (2nd ed., pp. 522–529). Lincoln, NE: Buros Institute of Mental Measurements.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, No. 1. Chicago: University of Chicago Press.

Timbrook, R. E., & Graham, J. R. (1994). Ethnic differences on the MMPI-2? *Psychological Assessment, 6*(3), 212217. doi: 10.1037/1040-3590.6.3.212

Timm, T., & Blow, A. (2005). The family life cycle and the genogram. In M. Cierpka, V. Thomas, & D. H. Sprenkle (Eds.), *Family assessment: Integrating multiple perspectives* (pp. 159–181). Cambridge, MA: Hogrefe.

Tindal, G. (2002). Large-scale assessments for all students: Issues and options. In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students*. Mahwah, NJ: Erlbaum.

Touliatos, J., Perlmutter, B. F., & Straus, M. A. (2001). *Handbook of family measurement techniques: Abstracts* (vol. I, II, III). Thousand Oaks, CA: Sage.

Tracey, T. J. G., & Hopkins, N. (2001). Correspondence of interests and abilities with occupational choice. *Journal of Counseling Psychology, 48*, 178–189. doi: 10.1037/0022-0167.48.2.178

Trull, T., & Widiger, T. A. (1997). *Structured Interview for the Five-Factor Model of Personality*. Odessa, FL: Psychological Assessment Resources.

Trzepacz, P. T., & Baker, R. W. (1993). *The psychiatric mental status examination*. New York: Oxford.

Tuddenham, R. D., Blumenkrantz, J., & Wilken, W. R. (1968). Age changes on AGCT: A longitudinal study of average adults. *Journal of Consulting and Clinical Psychology, 32*, 659–663.

Turner, S. M., DeMers, S. T., Fox, H. R., & Reed, G. (2001). APA's guidelines for test user qualifications: An executive summary. *American Psychologist, 56*, 1099–1113.

Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). Hoboken, NJ: Wiley.

U.S. Census Bureau. (2011). *National population by race*. Retrieved from http://www.census.gov/2010census/data/

U. S. Census Bureau. (2012). *Nearly 1 in 5 people have a disability in the U.S.* Retrieved from https://www.census.gov/newsroom/releases/archives/miscellaneous/cb12-134.html

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

U.S. Department of Defense. (2011). *ASVAB career exploration program: Theoretical and technical underpinnings of the revised skills composites and OCCU-Find*. Washington, DC: Author. Retrieved from http://www.asvabprogram.com/downloads/Technical_Chapter_2010.pdf

U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau. (2015). *Child maltreatment 2013*. Retrieved from http://www.acf.hhs.gov/programs/cb/research-data-technology/statistics-research/child-maltreatment

Vacc, N. A., & Juhnke, G. A. (1997). The use of structured clinical interviews for assessment in counseling. *Journal of Counseling and Development, 75*, 470–480.

Vacc, N. A., & Tippen, N. (2002). Documentation. In R. B. Ekstrom & D. K. Smith (Eds.), *Assessing individuals with disabilities in education, employment, and counseling settings* (pp. 59–70). Washington, DC: American Psychological Association.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*, 6–20.

Vacha-Haase, T., Tani, C. R., Kogan, L. R., Woodall, R. A., & Thompson, B. (2001). Reliability generalization: Exploring reliability variations on MMPI/MMPI-2 validity scale scores. *Assessment, 8*(4), 391–401. doi: 10.1177/107319110100800404

Van Brunt, B. (2015). *Harm to others: The assessment and treatment of dangerousness.* Alexandria, VA: American Counseling Association.

van der Linden, W. J., & Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 1–28). New York: Springer.

Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging, 18*, 332–339.

Vernon, P. E. (1950). *The structure of human abilities*. New York: Wiley.

Wang, L. (2002). Differential Aptitude Tests & Career Interest Inventory. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (4th ed., pp. 93–101). Tulsa, OK: National Career Development Association.

Warren, P. A. (2011). Effective psychological evaluation and management of behavioral health concerns. In P. A. Warren (Ed.), *Behavioral health disability: Innovation in prevention and management* (pp. 49–72). New York: Springer.

Warren, W. L. (1994). *Revised Hamilton Rating Scale for Depression*. Los Angeles: Western Psychological Services.

Watkins, C. E., Campbell, V. L., & Nieberding, R. (1994). The practice of vocational assessment by counseling psychologists. *The Counseling Psychologist*, *22*, 115–128.

Watson, J. C., & Flamez, B. (2015). *Counseling assessment and evaluation: Fundamental of applied practice.* Thousand Oaks, CA: Sage.

Watson, T. S., & Watson, T. S. (2009). Behavioral assessment in the schools. In A. Akin-Little & S. G. Little (Eds.), *Behavioral interventions in schools: Evidence-based positive strategies* (pp. 27–41). Washington, DC: American Psychological Association.

Wechsler, D. (1944). *The measurement of adult intelligence*. Baltimore: Williams & Wilkens.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2009). *Wechsler Individual Achievement Test* (3rd ed.). San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler Memory Scale-*(4th ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D., Kaplan, E., Fein, D., Kramer, J., Morris, R., Delis, D., & Maelender, A. (2004). *Wechsler Intelligence Scale for Children-Fourth Edition Integrated*. San Antonio, TX: The Psychological Corporation.

Wechsler, D., Raiford, S. E., & Holdnack, J. A. (2014). *Wechsler Intelligence Scale for Children-Fifth edition: Technical and interpretive manual.* Bloomington, MN: Pearson.

Weiner, I. B. (2013). Assessment of personality and psychopathology with performance-based measures. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology: Volume 2 testing and assessment in clinical and counseling psychology* (pp. 153–170). Washington, DC: American Psychological Association.

Weiss, D. J., Dawis, R. V., & Lofquist, L. H. (1981). *Minnesota Importance Questionnaire*. Minneapolis: Vocational Psychology Research.

Weiss, L. G., Saklofske, D. H., Prititera, A., & Holdnack, J. A. (2006). *WISC-IV advanced clinical interpretation*. San Diego: Academic Press.

Westbrook, B. W., Sanford, E., Gilleland, K., Fleenor, J., & Merwin, J. (1988). Career maturity in grade 9: The relationship between accuracy of self-appraisal and ability to appraise the career-relevant capabilities of others. *Journal of Vocational Behavior, 32*, 269–283.

Wheeler, D. L., Vassar, M., Worley, J. A., & Barnes, L. L. B. (2011). A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement, 71*(1), 231–244. doi: 10.1177/0013164410391579

Whipple, J. L., & Lambert, M. J. (2011). Outcome measures for practice. *Annual Review of Clinical Psychology, 7*, 87-111. doi: 10.1146/annurev-clinpsy-040510-143938

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: The use of early identified treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50*, 59–68.

Whiston, S. C. (1996). Accountability through action research: Research methods for practitioners. *Journal of Counseling & Development, 74*, 616–623.

Whiston, S. C. (2002). Application of the principles: Career counseling and interventions. *The Counseling Psychologist, 30*(2), 218–237. doi: 10.1177/0011000002302002

Whiston, S. C., Brecheisen, B. K., & Stephens, J. (2003). Does treatment modality affect career counseling effectiveness? *Journal of Vocational Behavior, 62*, 390–410.

Whiston, S. C., & Oliver, L. (2005). Career counseling process and outcome. In W. B. Walsh & M. Savickas (Eds.), *Handbook of vocational psychology* (3rd ed., pp. 155–194). Hillsdale NJ: Erlbaum.

Whiston, S. C., & Sexton, T. L. (1998). A review of school counseling outcome research: Implications for practice. *Journal of Counseling & Development, 76*, 412–426.

Whiston, S. C., Sexton, T. L., & Lasoff, D. L. (1998). Career-intervention outcome: A replication and extension of Oliver and Spokane (1988). *Journal of Counseling Psychology, 45*(2), 150–165. doi:10.1037/0022-0167.45.2.150

Whiston, S. C., Tai, W. L., Rahardja, D., & Eder, K. (2011). School counseling outcome: A meta-analytic examination of interventions. *Journal of Counseling & Development, 89*, 37–55. doi: 10.1002/j.1556-6678.2011.tb00059.x

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Wickwire, P. N. (2002). COP System. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (4th ed., pp. 212–217). Tulsa, OK: National Career Development Association.

Widiger, T. A. (1992). Review of the NEO Personality Inventory. In J. C. Conoley & J. J. Kramer (Eds.), *Eleventh mental measurements yearbook* (pp. 605–606). Lincoln, NE: The University of Nebraska Press.

Wiggins, J. S. (1966). Social desirability estimation and "faking good" well. *Educational and Psychological Measurement, 26*, 329–341.

Wilkinson, G. S., & Robertson, G. J. (2007). *Wide Range Achievement Test 4*. Lutz, FL: Psychological Assessment Resources.

Williams, S. K. (1978). The vocational card sort: A tool for vocational exploration. *The Vocational Guidance Quarterly, 48*, 237–243.

Willson, V. L., & Stone, E. (1994). Differential Aptitude Tests. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (pp. 90–98). Alexandria, VA: National Career Development Association.

Wilson, M. (2005). *Constructing measures: An item response model approach*. Mahwah, NJ: Erlbaum.

World Health Organization. (2012). *Are you ready? What you need to know about ageing.* Retrieved from http://www.who.int/world-health-day/2012/toolkit/background/en/

Worthington, E. L., McCullough, M. E., Shortz, J. L. Mindes, E. J., Sandage, S. J., & Chartrand, J. M. (1995). Can couples assessment and feedback improve relationships? Assessment as a brief relationship enrichment procedure. *Journal of Counseling Psychology, 42*, 466–475.

Wright, B. A., & Lopez, S. J. (2002). Widening the diagnostic focus: A case for including human strengths and environmental resources. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 26–44). New York: Oxford University Press.

Yesavage, J. A., Brink, T. L., Rose, T., Lum, O., Huang, O., Adey, V., & Leirer, V. (1983). Development and validation of the Geriatric Depression Screen Scale: A preliminary report. *Journal of Psychiatric Research, 17*, 37–49.

Zhang, D., Katsiyannis, A., Ju, S., & Roberts, E. (2014). Minority representation in special education: 5-year trends. *Journal of Child and Family Studies, 23*(1), 118–127. doi: 10.1007/s10826-012-9698-6

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Index

A

ABCL. *See* Adult Behavior Checklist
Abilities
  Campbell Interest and Skills Survey, 222–223
  self-estimates, 223–224
  Skills Confidence Inventory, 223
ACA. *See* American Counseling Association
Accessibility, 98
ACCI. *See* Adult Career Concerns Inventory
Accommodation, 98, 172
Acculturation, 107
Achenbach System of Empirically Based Assessment, 277–278
Achievement tests, 11
  adult, 199
  assessment of, 194–201
  diagnostic, 199–201
  high-stakes, 195–196
  individual, 199–201
ACT. *See* American College Testing
ACT Inc., 203–206
Acute stress disorder, 314
ADA. *See* Americans with Disabilities Act
ADAAA. *See* American with Disabilities Act Amendment Act
Adaptation, 98, 108–109
Adjustment disorder, 314
Adult achievement testing, 199
Adult Behavior Checklist (ABCL), 278
Adult Career Concerns Inventory (ACCI), 228
AERA. *See* American Educational Research Association
Affective assessment, 11
Affective Communication (AFC), 294
Affective instruments, 11
Age equivalent scores, 33–34
Aggression (AGG), 294
Agoraphobia, 313
Alternate form of reliability, 44
American Association of Marriage and Family Therapists, 293
American College Testing (ACT), 6
American Counseling Association (ACA), 3
American Counseling Association (ACA) *Code of Ethics*
  assessment security, 84
  conditions of assessment administration, 83
  description, 79–80
  forensic evaluation, 84–85
  general purpose, 80–81
  informed consent, 82
  instrument selection, 83
  level assessments, 81–82
  mental disorder diagnosis, 82–83
  multicultural issues/diversity in assessment, 83
  obsolete assessment and outdated results, 84
  qualified personnel, 82
  scoring and interpretation, 83–84
American Educational Research Association (AERA), 3, 15, 196
American Psychiatric Association, 301
American Psychological Association (APA), 3, 15, 196
Americans with Disabilities Act (ADA), 89
American with Disabilities Act Amendment Act (ADAAA), 89–90
Analogue observation, 269–270
Analysis, data, 340
Anorexia nervosa, 315
Antisocial personality disorder, 321
Anxiety disorders, 312–313
APA. *See* American Psychological Association
Appearance, 165
Appraisal, 3
Aptitude tests, 11
  career, 207–211
  differential, 210–211
  scholastic, 201–207
  vocational, 207–211
Arithmetic concept, 178

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Armed Services Vocational Aptitude Battery (ASVAB), 29, 208–210
Arrangement techniques, 262–263
Assessment
  affective, 11
  authentic, 128–129
  cognitive, 11
  in counseling, 4–17
  counselors, knowledge of, 3–4
  couples, 294–296
  culture and, 98–109
  definitions, 2–3
  depression, 158–159
  diversity, 83
  family, 282–298
  group, 10
  history of, 11–17
  individual, 10, 291–298
  informal techniques, 238–240
  informed consent, 82
  legal issues, 87–94
  marital, 292–298
  nonstandardized, 9
  nonverbal, 10
  objective, 10
  performance, 128–129
  potential violence, 159–160
  power, 10
  premarital, 293–294
  psychological reports, 137
  psychological test, 2
  scoring, 10
  scoring and interpretation, 83–84
  search resources, 121–122
  security, 84
  selection of, 335–339
  speed, 10
  standardized, 9
  strategies, 7
  subjective, 10
  therapeutic model, 7
  types of tools, 9–11
  verbal, 10
Assessment for Signal Clients, 331
Assessment instruments
  administering, 126
  determining necessary information, 120
  evaluate strategies, 122–125
  results from, communication of, 130–138
  scoring systems, 126–129
  search assessment resources, 121–122
  selection of, 120–125
  strategies for obtaining information, 120–121
Assimilation, 172
ASVAB. *See* Armed Services Vocational Aptitude Battery
Attention-deficit/hyperactivity disorder (ADHD), 303, 305, 309–310
Attitude, 165
Authentic assessment, 128–129
Autism spectrum disorder, 309
Avoidant personality disorder, 322
Avoidant/restrictive food intake disorder, 315

B

BASC-2. *See* Behavior Assessment System for Children, Second Edition
Base rates, 325
BDI. *See* Beck Depression Inventory
BDI-II. *See* Beck Depression Inventory-II
Beck Depression Inventory (BDI), 159, 291, 337
Beck Depression Inventory-II (BDI-II), 329, 336–337
Behavioral assessment
  analogue observation, 269–270
  definition, 268
  error in, 278–279
  formal, 277–278
  naturalistic observation, 269
  target behavior, 268
Behavioral observations, 136
Behavior Assessment System for Children, Second Edition (BASC-2), 277
Bias
  content, 99–100
  intercept, 102
  observer, 279
  slope, 101
  tests, 93–94, 99
Binge-eating disorder, 316
Bipolar disorder, 311
Block Design, 177–178
Body dysmorphic disorder, 314
Borderline personality disorder, 321
Brief Symptoms Inventory (BSI), 332
Brief Symptoms Inventory 18 (BSI-18), 332
BSI. *See* Brief Symptoms Inventory
BSI-18. *See* Brief Symptoms Inventory 18
Bulimia nervosa, 316

C

CAAS. *See* Career Adapt-Abilities Scale
CACREP. *See* Council for Accreditation of Counseling and Related Educational Programs
CAI. *See* Career Assessment Inventory
Campbell Interest and Skills Survey (CISS), 222–223
Cannot Say Score (?) Scale, 244
Career Adapt-Abilities Scale (CAAS), 229, 338
Career adaptability, 228–229
Career aptitude tests, 207–211
Career assessment
  ethnic and cultural differences, 234–235
  gender and, 233–234
  Internet-based, 231–233
  issues and trends, 231–235
  multicultural, 234–235
  qualitative, 229–231
  technology-based, 231–233
Career Assessment Inventory (CAI), 219–220
Career counseling, 7, 337–338
Career decision making, 227–228

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Career Decision-Making Difficulties Questionnaire (CDDQ), 227–228
Career Decision-Making Self-Efficacy Scale (CDMSE), 227
Career Decision Scale (CDS), 227
Career Development Inventory (CDI), 228
Career maturity, 228
Career Maturity Inventory, 338
Career process measures, 227–229
Career Thoughts Inventory, 229
Career Transitions Inventory, 229
CAS. *See* Counseling Aptitude Scale
CAT. *See* Computer-adaptive testing
Cattell-Horn-Carroll model, 171
CBCL/6-18. *See* Child Behavior Checklist System/6-18
CDDQ. *See* Career Decision-Making Difficulties Questionnaire
CDI. *See* Career Development Inventory
CDMSE. *See* Career Decision-Making Self-Efficacy Scale
CDS. *See* Career Decision Scale
Centers for Disease Control and Prevention, 153
Central tendency, measures of, 23–25
Chaotic Scale, 296
Child abuse, 160–162
Child Behavior Checklist System/6-18 (CBCL/6-18), 278
Childhood-onset fluency disorder (stuttering), 309
Child Protective Services (CPS), 160
Children interviewing, 146–147
Circumplex Model, 283, 296
CISS. *See* Campbell Interest and Skills Survey
Civil Rights Act, 88–89
Classical test theory, 38–39
Client rights and responsibilities, 85–86
Client Satisfaction Questionnaire, 337
Client's problems
  assessment of, 153–166
  child abuse, 160–162
  defining, 144–145
  depression, 158–159
  mental status examination, 164–166
  substance abuse, 162–164
  suicide, potential of, 153–158
  violence, potential, 159–160
Clinical Personality Patterns, 325
Cluster sampling, 35
Coaching, 211–212
Code of Fair Testing Practices in Education, 120
Coefficient alpha, 46
Coefficient of determination, 42
CogAT. *See* Cognitive Abilities Test
Cognition, 165
Cognitive Abilities Test (CogAT), 185
Cognitive assessment, 11
Cognitive impairments, 113–115
College Entrance Examination Board (CEEB) scores, 33
Common Core State Standards, 197–198
Communicating results
  to clients, 131–132
  to parents, 133–134
  psychological reports, 134–138
  research related to, 130–131
Competency testing, 94
Completion techniques, 262
Composition, 285
Comprehension subtest, 177
Computer adaptive testing, 76
Computer-adaptive testing (CAT), 198
Computer-assisted career assessment systems, 232
Concurrent validity, 64
Conditional standard errors of measurement, 53
Conduct disorder, 307, 317
Construction techniques, 261
Construct-irrelevance validity, 57
Construct underrepresentation validity, 57
Construct validity, 58
Content bias, 99–100
Content-oriented validation evidence, 60
Content-related validity, 57–58
Convergent evidence, 63
Conversion disorder, 315
Conversion Hysteria Scale, 246
COPSystem, 226
Correction (K) Scale, 245
Correlation
  definition, 39–40
  scatterplots for, 40
Correlational item discrimination method, 73
Correlational validity method, 62
Correlation coefficient
  definition, 40–41
  Pearson-Product Moment, 41–43
Council for Accreditation of Counseling and Related Educational Programs (CACREP), 3, 8
Counseling
  accountability of, 339–340
  assessments during, 4–17
  client's problems, 4–5
  conditions, 323
  evaluation in, 333–340
  evaluation of, 6–7
  integral assessment, 4–7
  school, 338–339
  selecting and implementing treatment, 5–6
  treatments during, 328–333
Counseling Aptitude Scale (CAS), 18, 21, 67
Counselors
  culturally competent professional, 103–105
  implications for, 324
  knowledge of, 3–4, 8–9
  rights and responsibilities, 86–87
CPS. *See* Child Protective Services
Criterion-referenced instruments
  definition, 20–21
  norm-referenced instrument *vs.*, 20–21

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Criterion-related validity, 58
Cronbach's Alpha, 46
Cross-sectional studies, 186
Crystallized abilities, 171
Cultural Formulation Interview (CFI), 304
Culturally competent professional counselors, 103–105
Cultural specificity, 234
Cultural validity, 234
Culture, 98–109
Cyclothymic disorder, 311

D

DAS-II. *See* Differential Ability Scales-Second Edition
DAT. *See* Differential aptitude test
Data analysis, 340
Debra P. v. Turlington (1981), 94
Decision theory, 67–69
Delusional disorder, 310
Dependent personality disorder, 322
Depersonalization/derealization disorder, 315
Depression
  assessment, 158–159
  elevated scores, 246
  instruments, 159
Depressive disorder, 312
Descriptive assessments, 240
Design, 334
Developmental coordination disorder, 310
Developmental progressions, 171–172
Deviation IQs, 33
Diagnosis of disorders
  DSM-5, 301
  DSM-IV-TR, 317–319
  instruments designed for, 324–326
Diagnostic achievement testing, 199–201
Diagnostic and Statistical Manual of Mental Disorders (DSM), 16
Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-5), 301–303
  cultural issues, 304
  definition of mental disorder, 302–303
  emerging measures, 306–307
  nonaxial system, 304–305
  organizational structure, 303
Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV), 302
Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition-Text Revision's (DSM-IV-TR), 307, 309, 318, 319, 326
Diagnostic assessments, 240
DIF. *See* differential item functioning
Differential Ability Scales—Second Edition (DAS-II), 184
Differential aptitude test (DAT), 210–211
Differential item functioning (DIF), 62, 100
Digit Span, 178
Direct behavior, 268
Disabilities
  assessment of, within diverse populations, 109–117
  cognitive impairments, 113–115
  hearing impairments, 112
  motor disabilities, 112–113
  older individuals, 115–117
  visual impairments, 111–112
Disagreement About Finances (FIN), 294
Discriminant evidence, 63
Disinhibited social engagement disorder, 314
Disorders
  acute stress, 314
  addictive, 317–318
  adjustment, 314
  antisocial personality, 321
  anxiety, 312–313
  autism spectrum disorder, 309
  avoidant personality, 322
  avoidant/restrictive food intake, 315
  bipolar, 311
  bipolar I, 311
  bipolar II, 311
  body dysmorphic, 314
  borderline personality, 321
  childhood-onset fluency, 309
  conduct, 307, 317
  conversion, 315
  cyclothymic, 311
  delusional, 310
  dependent personality, 322
  depressive, 312
  developmental coordination, 310
Diagnostic and Statistical Manual of Mental, 301–302, 301–303
  disruptive, 317
  dissociative, 315
  eating, 315–316
  exhibitionistic, 322
  factitious, 315
  feeding, 315–316
  fetishistic, 323
  frotteuristic, 323
  gambling, 318
  histrionic personality, 321
  impulse-control, 317
  intellectual developmental, 309
  intermittent explosive, 317
  major depressive, 312
  mental, 302–303
  motor, 310
  narcissistic personality, 321–322
  neurocognitive, 318–319
  neurodevelopmental, 309–310
  obsessive-compulsive, 313–314
  obsessive-compulsive personality, 322
  oppositional defiant, 317
  other mental, 323
  paranoid personality, 320
  paraphilic, 322–323
  pedophilic, 323
  persistent depressive, 312
  Pica, 315
  premenstrual dysphoric, 312
  psychotic, 310
  reactive attachment, 314
  rumination, 315
  schizoid personality, 320
  schizotypal personality, 320–321

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

separate anxiety, 312–313
sexual masochism, 323
social anxiety, 313
social (pragmatic) communication, 309
somatic symptom, 315
specific learning disorder, 310
specified and unspecified, 305–307
speech sound, 309
stereotypic movement, 310
substance use, 317–318
Tourette's, 310
transvestic, 323
voyeuristic, 322
Disruptive disorder, 317
Disruptive mood dysregulation disorder, 312
Dissatisfaction with Children (DSC), 294
Disseminating results, 340
Dissociative amnesia, 315
Dissociative disorder, 315
Dissociative identity disorder, 315
Distributions
frequency, 22
negatively skewed, 28
normal, 27–30
positively skewed, 28
skewed, 28
Diverse populations
bias testing, 99
culture and, 98–109
individual disabilities assessments, 109–117
linguistic background and appraisal, 107–109
performance and race/ethnicity, 105–107
recommendations for practice, 102–105
Diversity of family, 285
Domain sampling theory, 52–53
Dominant factor, 74
Draw-a-Person (D-A-P) test, 263
Drawing techniques, 263
DSM. *See* Diagnostic and Statistical Manual of Mental Disorders
DSM-5, 302–303
Duration, 268–269
Duty to warn, 160
Dyadic Adjustment, 294
Dyadic Adjustment Scale (DAS), 294
Dyadic Relationship Scale, 296
Dysfunctional Attitude Scale, 337

E

Eating disorder, 315–316
Ecomap, 291
Educational Testing Service (ETS), 203
Educational Testing Service (ETS) Test Link, 122
EEOC. *See* Equal Employment Opportunity Commission
Elementary and Secondary Education Act (ESEA), 195
Empirical criterion keying, 241
Equal Employment Opportunity Commission (EEOC), 88
Error
behavioral assessment, 278–279
random, 39
systematic, 39
unsystematic, 39
ESEA. *See* Elementary and Secondary Education Act
Ethical Principles of Psychologists *and Code of Conduct*, 79
Ethics. *See* American Counseling Association (ACA) Code of Ethics
Ethnic group differences, 190–191
Etic view, 99
Evaluation
in counseling, 333–340
formative, 333
summative, 333
Evaluation study focus, defining, 333–334
Evidence
consequences of testing, 70
convergent, 63
discriminant, 63
instrument content, 59–60
internal structure, 61–62
relations to other variables, 62–70
response processes, 60–61
Excoriation, 314
Exhibitionistic disorder, 322
Expectancy tables, 69
Exploration, 144
Expression techniques, 263–264
Extreme group method, 72–73

F

face validity, 60
Factitious disorder, 315
Factor analysis, 61–62, 241
Fairness, 123
False negative, 69
False positive, 69
Families assessment, 283
difficulties with, 283–284
individual assessment, 291–292
instruments for, 296–298
interviews with, 285–286
marriage and, formalized instruments for, 292–298
Family Adaptability and Cohesion Evaluation Scale IV (FACES IV), 296
Family affect, 285
Family Assessment Measure (FAM-III), 296–297
Family diversity, 285
Family Educational Rights and Privacy Act (FERPA), 15, 91–92
Family Environment Scale (FES), 297–298
Family History of Distress (FAM), 294
Family organization, 285
Family Problem Scale (FAM), 291
Family processes, 285
Family sculpting, 289–290
Family structure, 285
Feeding disorder, 315–316
FERPA. *See* Family Educational Rights and Privacy Act
Fetal alcohol syndrome, 189
Fetishistic disorder, 323
Figure Weights, 178
Fluid abilities, 171

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Fluid reasoning subtests, 178
Flynn effect, 191
Formal behavioral assessment
  Achenbach System of Empirically Based Assessment, 277–278
  Behavior Assessment System for Children, Second Edition, 277
Formative evaluation, 333
Frequency, 268
Frequency distribution, 22
Frequency polygon, 22
Frotteuristic disorder, 323
Functional neurological disorder, 315

G

Galton, Francis, 11
Gambling disorder, 318
GATB. *See* General Aptitude Test Battery
GED. *See* General Education Development
Gender
  career assessment, 233–234
  intelligence testing, 189–190
General ability factor, 169
General ability tests, 11
General Aptitude Test Battery (GATB), 187
General Education Development (GED), 199
Generalizability theory, 52–53
General Scale, 296
Genograms, 286–289
Georgia NAACP v. State of Georgia (1985), 94
GI Forum v. Texas Education Agency (1997), 94
Global Assessment of Functioning (GAF), 307
Global developmental delay, 309
Global Distress Scale (GDS), 294
Goal Attainment Scaling (GAS), 332, 338
Grade equivalent norms, 33
Grade equivalent scores, 33–34
Graduate Record Examination (GRE), 206
GRE. *See* Graduate Record Examination
Griggs v. Duke Power Company (1971), 88
Grossman Facet Scales, 325
Group assessment, 10
Group intelligence testing, 184–185
Group separation, 67

H

Hamilton Rating Scale for Depression (HRSD), 337
Hassles Scale, 337
Health Insurance Portability and Accountability Act (HIPPA), 92–93
Hearing impairments, 112
Hereditary Genius (Galton), 12
Hereditary intelligence testing, 188
Heritability index, 188
High-stakes testing, 195–196
HIPPA. *See* Health Insurance Portability and Accountability Act
Histogram, 23
Histrionic personality disorder, 321
House-Tree-Person (H-T-P) test, 263
Hypochondriasis Scale, 246
Hypomania Scale, 247
Hypomanic episode, 311

I

IDEA. *See* Individuals with Disabilities Education Act
IEP. *See* Individualized Education Plan
Illness anxiety disorder, 315
Impulse-control disorder, 317
Indirect behavior, 268
Individual achievement testing, 199–201
Individual assessments, 10, 291–292
  cognitive impairments, 113–115
  within diverse populations, 109–117
  hearing impairments, 112
  motor disabilities, 112–113
  older individuals, 115–117
  visual impairments, 111–112
Individual intelligence testing
  additional individual instruments, 184
  Kaufman instruments, 182–183
  Stanford-Binet intelligence scale, 180–182
  Wechsler scales, 174–177
  Woodcock-Johnson IV, 183
Individualized Education Plan (IEP), 90–91
Individuals with Disabilities Education Act (IDEA), 90–91
Informal techniques for assessment
  interviewing, 240
  observation, 239–240
  for personality assessment, 238–240
Information-gathering model, 7
Information processing intelligence model, 172–174
Information subtest, 177
Informed consent, 82
Infrequency Back [F(B)] Scale, 244
Infrequency (F) Scale, 244
Initial counseling assessment
  checklists, 150–152
  client's problem, defining, 144–145
  initial intake interview, 141–143
  interviewing, 145–150
  problems, assessment of specific, 153–166
  rating scales, 152–153
Initial intake interview
  client background information, 142
  demographic information, 142
  health and medical history, 142–143
  information gathered, 142
  other relevant information, 144
  presenting concerns of client, 143
Insight, 165
Instruments
  accessibility, 98
  accommodation, 98
  adaptation, 98, 108–109
  bias testing, 99
  content bias, 99–100
  criterion relationships, 100–102
  depression, 159

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

for diagnosis of disorders, designed for, 324–326
for families assessment, 296–298
internal structure, 100
Million Clinical Multiaxial, 325–326
modification, 98, 108–109
slope bias, 101
substance abuse, 164
suicide potential, 157–158
testing *vs*., 3
translation, 108–109
universal design, 98
use and interpretation, 109
Integrative career assessment programs
COPSystem, 226
Kuder Career Planning System, 225
Intellectual disability, 309
Intelligence models
Cattell-Horn-Carroll model, 171
developmental progressions, 171–172
information processing, 172–174
other theories, 174
psychometric approach, 169–171
Intelligence quotients (IQs)
definition, 12
deviation, 33
Intelligence tests, 11
environmental factors, 188–189
ethnic group differences, 190–191
Flynn effect, 191
gender differences, 189–190
group, 184–185
hereditary, 188
individual, 174–184
issues in, 186–191
scoring predictions, 187–188
stability, 186
Intensity, 269
Interaction patterns, 283
Intercept bias, 102
Intercepts, 102
Interest
Career Assessment Inventory, 219–220
Jackson Vocational Interest Survey, 221
Judgment of Occupational Behavior-Orientation, 221
O*NET Interest Profiler, 220–221
Self-Directed Search, 220
Strong Interest Inventory, 216–219
Wide Range Interest-Opinion Test-Second Edition, 221
Intermittent explosive disorder, 317
Internal consistency measures of reliability, 44–46
International Classification of Diseases (ICD), 301
International Classification of Diseases-Tenth Edition (ICD-10), 301–302
International Society for Sandplay Therapy, 262
Internet-based career assessment, 231–233
Interpersonal functioning, 337
Interrater reliability, 47
Intersubject design, 334
Interval sampling, 271
Interval scale, 19–20
Interval time sampling, 271
Interviewing
children, 146–147
during families assessment, 285–286
informal techniques for assessment, 240
limitations, 149–150
psychosocial, 142
semistructured, 147
skills, 145
strengths, 149–150
structured, 147–149
techniques, 145
unstructured, 147–149
Intrasubject design, 334
Inventories
Beck Depression, 291, 337
Career Maturity, 338
Million Multiaxial, 291
Quality of Life, 337
State-Trait Anxiety, 337
In View, 185
Ipsative interpretation, 224
IQs. *See* Intelligence quotients
Item analysis
definition, 71
item difficulty, 71–72
item discrimination, 72–73
item response theory, 73–77
Item characteristic curve, 74
Item characteristic function, 74
Item difficulty, 71–72
Item discrimination, 72–73
Item response theory (IRT), 73–77
age/grade equivalent scores, 33
reliability, 53–54

J

Jackson Vocational Interest Survey (JVIS), 221
JOB-O. *See* Judgment of Occupational Behavior-Orientation
Judgment, 165
Judgment of Occupational Behavior-Orientation (JOB-O), 221
JVIS. *See* Jackson Vocational Interest Survey

K

Kaufman instruments, 182–183
Kaufman Test of Educational Achievement, Third Edition (KTEA-3), 200
KCIA. *See* Kuder® Career Interests Assessment™
KCPS. *See* Kuder Career Planning System
Kinetic Family Drawing (K-F-D) test, 263
Kleptomania, 317
Knowledge
acquisition, 173
counselors, 3–4, 8–9
KSCA. *See* Kuder® Skills Confidence Assessment™
KTEA-3. *See* Kaufman Test of Educational Achievement, Third Edition
Kuder® Career Interests Assessment™ (KCIA), 225

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Kuder Career Planning System (KCPS), 225
Kuder–Richardson formulas, 45–46
Kuder® Skills Confidence Assessment™ (KSCA), 225
Kuder® Work Values Assessment™ (KWVA), 225
Kvebaek Family Sculpture Technique, 290
KWVA. *See* Kuder® Work Values Assessment™

L

Language analysis, 165
Language disorder, 309
Larry P. v. Riles (1979), 93
Latent-trait theory, 54
Legislation
  American with Disabilities Act Amendment Act, 89–90
  Civil Rights Act, 88–89
  Family Educational Rights and Privacy Act, 91–92
  Health Insurance Portability and Accountability Act, 92–93
  Individuals with Disabilities Education Act, 90–91
Letter-numbering sequencing, 178
Lie (L) Scale, 245
Litigation
  minimum competency, 94
  test bias and placement, 93–94
Local independence, 74
Longitudinal studies, 186

M

MAB-II. *See* Multidimensional Aptitude Battery II
Major depressive disorder, 312
Major depressive episode, 311
Major neurocognitive disorders, 318
Manic episode, 311
Mapping techniques, 286–291
Marital Adjustment Test, 294
Marital Satisfaction Inventory-Revised (MSI-R), 294–295
Marriage
  couples assessment, 294–296
  marital assessment, 292–298
  premarital counseling, 293–294
Marriage Study Consortium at Brigham Young University, 293
Masculinity-Femininity Scale, 247
Matrix Reasoning, 178
MCMI-III. *See* Millon Clinical Multiaxial Inventory-III (MCMI-III)
Mean, 25
Measurement
  career process, 227–229
  internal consistency, 44–46
  in mental health setting, 337
  in school counseling, 338–339
  selection of, 335–339
  self-concept, 265
  standard error of, 48–51
Measurement scales
  description, 19
  interval scale, 19–20
  nominal scale, 19
  ordinal scale, 19
  ratio scale, 20
Median, 24–25
Mental disorder, 302–303, 323
  diagnosis, 82–83
Mental factors, 169
Mental health setting, 337
Mental Measurements Yearbook (MMY), 121
Mental operations, 170
Mental retardation, 309
Mental status examination (MSE), 164–166
Metacomponents, 173
MI. *See* Motivational interviewing
Mild neurocognitive disorders, 318
Millon Adolescent Clinical Inventory (MACI), 325–326
Millon Clinical Multiaxial Inventory-III (MCMI-III), 325–326
Millon Index of Personality Styles (MIPS), 326
Millon Multiaxial Inventory, 291
Minnesota Importance Questionnaire (MIQ), 224
Minnesota Multiphasic Personality Inventory-2 (MMPI-2), 11, 15, 291, 298, 329
  Cannot Say Score (?) Scale, 244
  Conversion Hysteria Scale, 246
  Correction (K) Scale, 245
  depression elevated scores, 246
  Hypochondriasis Scale, 246
  Hypomania Scale, 247
  Infrequency Back [F(B)] Scale, 244
  Infrequency (F) Scale, 244
  Lie (L) Scale, 245
  Masculinity-Femininity Scale, 247
  Paranoia Scale, 247
  Psychasthenia Scale, 247
  psychometric qualities, 249–250
  Psychopathic Deviate Scale, 246–247
  Psychopathology Infrequency [F(p)] Scale, 244–245
  Schizophrenia Scale, 247
  Social Introversion Scale, 248
  Superlative Self-Presentation (S) Scale, 245
  supplementary scales, 248
  Symptom Validity (FBS) Scale, 245
  three-point codes, 248
  True Response Inconsistency (TRIN) Scale, 244
  two-point codes, 248
  validity scales, 244
  Variable Response Inconsistency (VRIN) Scale, 244
MIPS. *See* Millon Index of Personality Styles (MIPS)
MIQ. *See* Minnesota Importance Questionnaire
MMPI-2. *See* Minnesota Multiphasic Personality Inventory-2
MMPI-2 Restructured Form, 250–252
MMY. *See* Mental Measurements Yearbook
Mode, 23
Models
  Circumplex, 283, 296
  Olson Circumplex, 283

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

one-parameter, 75
Rasch, 75
three-parameter, 75–76
two-parameter, 75
Modification, 98, 108–109
Monitoring of treatments during counseling, 329–333
Mood and affect, 165
Mood episodes, 311
Moos, Rudolf, 297
Motivational interviewing (MI), 163–164
Motor disabilities, 112–113
Motor disorders, 310
MSE. *See* Mental status examination
Multicultural career assessment, 234–235
Multidimensional Aptitude Battery II (MAB-II), 185
Multiple regression, 66
Multitrait-multimethod matrix, 63
Myers-Briggs Type Indicator® (MBTI), 137, 255–259, 282, 291, 298

N

NAEP. *See* National Assessment of Educational Progress
Narcissistic personality disorder, 321–322
National Assessment of Educational Progress (NAEP), 198–199
National Career Development Association (NCDA), 232
National Council on Measurement in Education (NCME), 3, 15
National Counselor Examination, 3
National Institute on Drug Abuse, 296
Naturalistic observation, 269
NCDA. *See* National Career Development Association
NCME. *See* National Council on Measurement in Education
Negatively skewed distributions, 28
NEO Personality Inventory-3 (NEO PI-3), 252–255
Neurodevelopmental disorder, 309–310
No Child Left Behind (NCLB) Act, 16, 195–196
Nominal scale, 19
Nonaxial system, 304–305
Nonstandardized assessment, 9
Nonverbal assessment, 10
Normal curve, 27. *See also* Normal distributions
Normal distributions, 27–30
Norming group
evaluation, 35–36
selection of, 122–123
Norm-referenced instruments
criterion-referenced instruments *vs.*, 20–21
definition, 20
interpretation, 21–23

O

Objective assessment, 10
Observation, 239–240
analogue, 269–270
naturalistic, 269
Observational recording methods
event recordings, 273
interval recordings, 271–273
narrative recordings, 270–271
ratings recordings, 274–275
Observer bias, 279
Observer drift, 279
Obsessive-compulsive disorder, 313–314
Obsessive-compulsive personality disorder, 322
Older individuals, 115–117
OLSAT-8. *See* Otis-Lennon School Ability Test, Eighth Edition
Olson Circumplex Model, 283
One-parameter model, 75
O*NET Interest Profiler, 220–221
Opportunity dominance, 233–234
Oppositional defiant disorder, 317
Ordinal scale, 19
Organizational structure, 303
Otis-Lennon School Ability Test, Eighth Edition (OLSAT-8), 185
Outcome Questionnaire (OQ), 329, 331, 337

P

Panic attack, 313
Panic disorder, 313
Parallel form of reliability, 44
Paranoia Scale, 247
Paranoid personality disorder, 320
Paraphilic disorders, 322–323
PARCC. *See* Partnership for Assessment of Readiness for College and Careers
Parent Rating Scale (PRS), 277
Parents in Action on Special Education (PASE) v. Hannon (1980), 93
Partial-interval time sampling, 272
Participants, selection of, 334
Partners for Change Outcome Management System (PCOMS), 329–330
Partnership for Assessment of Readiness for College and Careers (PARCC), 197
Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4), 184
Pearson-Product Moment correlation coefficient, 41–43
Pedophilic disorder, 323
Percentile ranks, 29
Percentile score, 29
Perception, 165
Performance assessment, 128–129
Performance-based measures, 238
Performance components, 173
Performance tests, 10
Persistent depressive disorder, 312
Personality disorder, 319
antisocial, 321
avoidant, 322
borderline, 321
dependent, 322
histrionic, 321
narcissistic, 321–322
paranoid, 320
schizoid, 320
schizotypal, 320–321

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Pica, 315
Picture Concept, 178
Picture Span, 178
Piers-Harris Children Self-Concept Scale, Second Edition, 265
Positively skewed distributions, 28
Posttraumatic stress disorder (PTSD), 314
Power assessment, 10
PPVT-4. *See* Peabody Picture Vocabulary Test, Fourth Edition
Predictive validity, 64
Premarital counseling, 293–294
PREmarital Personal and Relationship Evaluation (PREPARE), 293–294
Premenstrual dysphoric disorder, 312
Primary mental abilities, 170
Problem-Solving Communication (PSC), 294
Processing speed subtests, 179–180
Products, 333
Progressive matrices, 184
Projective techniques, 11
  arrangement techniques, 262–263
  association techniques, 260–261
  completion techniques, 262
  construction techniques, 261
  definition, 238
  expression techniques, 263–264
  limitations, 264–265
  selection techniques, 262–263
  strengths, 264–265
PRS. *See* Parent Rating Scale
Psychasthenia Scale, 247
Psychological reports
  assessment results and interpretations, 137
  background information, 136
  behavioral observations, 136
  common areas, 135
  definition, 134
  diagnostic impressions and summary, 137–138
  identifying information, 135
  reason for referral, 135–136
  signature, 138
Psychological test, 2
Psychometric intelligence model, 169–171
Psychometric qualities, 249–250
Psychopathic Deviate Scale, 246–247
Psychopathology Infrequency [F(p)] Scale, 244–245
Psychosocial interview, 142
Psychotherapy Readiness Scale, 330
Psychotic disorders, 310
Pyromania, 317

Q

Qualitative career assessment, 229–231
Qualitative data, 334
Quality of Life Inventory, 337
Quantitative data, 334
Questionnaires
  Client Satisfaction, 337
  Outcome, 329, 331, 337

R

Random error, 39
Randomized clinical trials design, 334
Range, 25
Rasch model, 75
Rate, 268
Rating scales, 152–153
Ratio scale, 20
Raw score, 29
Reactive attachment disorder, 314
Reactivity, 276–277
Regression
  equation, 65
  line, 65
  validity, 64–67
RELATE, 293–294
Relationship patterns, 285
Relations to other variables
  decision theory, 67–69
  prediction/instrument-criterion relationships, 63–64
  regression, 64–67
  validity generalization, 70
Reliability
  alternate forms, 44
  assessment instruments, 123
  classical test theory, 38–39
  coefficient alpha, 46
  conditional standard errors of measurement, 53
  definition, 37
  domain sampling theory, 52–53
  evaluating coefficients, 47–48
  generalizability theory, 52–53
  instruments, 46–47
  internal consistency measures, 44–46
  interrater, 47
  item response theory, 53–54
  Kuder–Richardson formulas, 45–46
  nontypical situations, 46–47
  parallel forms, 44
  random error and, 39
  Spearman–Brown formula, 45
  split-half, 45
  standard error of difference for, 51
  standard error of measurement, 48–51
  test-retest, 43–44
  universe score, 53
Response-to-intervention (RTI), 200
Responsibilities of Users of Standardized Tests (RUST) statement, 79–80
Rights and responsibilities
  client, 85–86
  counselors, 86–87
Rigid Scale, 296
Role Orientation (ROR), 294
Rorschach Inkblot Test, 260–261
Rotter Incomplete Sentences Blanks, Second Edition, 262
RTI. *See* Response-to-intervention
Rumination disorder, 315

S

Sample/sampling
  cluster, 35
  simple random, 35
  stratified, 35
SAT. *See* Scholastic aptitude tests
Scale
  Cannot Say Score (?), 244

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Career Adapt-Abilities, 338
Chaotic, 296
Conversion Hysteria, 246
Correction, 245
Dyadic Relationship, 296
Dysfunctional Attitude, 337
Family Environment, 297–298
Family Problem, 291
General, 296
Goal Attainment, 332, 338
Grossman Facet, 325
Hassles, 337
Hypochondriasis, 246
Hypomania, 247
Infrequency, 244
Infrequency Back, 244
interval, 19–20
Lie, 245
Masculinity-Femininity, 247
measurement, 19–20
nominal, 19
ordinal, 19
Paranoia, 247
Piers-Harris Children Self-Concept Scale, Second Edition, 265
Psychasthenia, 247
Psychopathic Deviate, 246–247
Psychopathology Infrequency, 244–245
Psychotherapy Readiness, 330
rating, 152–153
ratio, 20
Rigid, 296
Schizophrenia, 247
Self-Rating, 296
Severe Personality Pathology, 325
Social Introversion, 248
Stanford–Binet, 12
suicide probability, 157–158
Superlative Self-Presentation, 245
supplementary, 248
Symptom Validity, 245
Teacher Rating, 277
Tennessee self-concept, 265
Therapeutic Bond, 330
True Response Inconsistency, 244
validity, 244, 294, 325
Variable Response Inconsistency, 244
Schedule for Affective Disorders and Schizophrenia, 337
Schizoaffective disorder, 310
Schizoid personality disorder, 320
Schizophrenia, 310
Schizophrenia Scale, 247
Schizophreniform disorder, 310
Schizotypal personality disorder, 320–321
Scholastic aptitude tests (SAT)
ACT Inc., 203–206
Graduate Record Examination, 206
validity, 207
Scholastic Assessment Test, 202
School counseling, 338–339
SCI. *See* Skills Confidence Inventory
SCICA. *See* Semistructured Clinical Interview of Children & Adolescents
SCID-5. *See* Structured Clinical Interview for DSM-5 (SCID-5)
SCID-5-CV (Clinician Version), 324
SCID-5-RV (Research Version), 324
Scores
age equivalent, 33–34
assessment tools, 10
CEEB, 33
depression elevated, 246
deviation IQs, 33
grade equivalent, 33–34
percentile, 29
raw, 29
standard, 30–34, 332
stanines, 32
T, 31–32
types of, 28–30
z, 31
Scoring systems, 126–129
SDH. *See* Structured Developmental History
SDS. *See* Self-Directed Search
SEA. *See* State education agencies
Search assessment resources, 121–122
Selection techniques, 262–263
Selective mutism, 313
Self-concept measures, 265
Self-Directed Search (SDS), 220
Self-estimates of abilities, 223–224
Self-monitoring, 275–276
Self-Rating Scale, 296
Self-Report of Personality (SRP), 277
Semistructured Clinical Interview of Children & Adolescents (SCICA), 278
Semistructured interviewing, 147
Separate anxiety disorders, 312–313
Sequential processing, 172
Severe Personality Pathology Scales, 325
Sexual Dissatisfaction (SEX), 294
Sexual masochism disorder, 323
Signature, 138
Similarities subtest, 177
Simple random sample, 35
Simultaneous processing, 172
SIT-R3. *See* Slosson Intelligence Test-Revised, Third Edition
Skewed distributions, 28
Skills Confidence Inventory (SCI), 223
Slope bias, 101
Slosson Intelligence Test-Revised, Third Edition (SIT-R3), 184
Smarter Balanced Assessment Consortium, 197
Social anxiety disorder, 313
Social (pragmatic) communication disorder, 309
Social Introversion Scale, 248
Socialization dominance, 233–234
Somatic symptom disorder, 315
Spearman–Brown formula, 45
Spearman's intelligence model, 169
Specific learning disorder, 310
Specific phobia, 313
Specified disorder, 305–307
Speech analysis, 165
Speech sound disorder, 309
Speed assessment, 10
Split-half reliability, 45
SRP. *See* Self-Report of Personality
Standard deviation, 26
Standard error of difference, 51

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Standard error of estimate, 66–67
Standard error of measurement, 48–51
Standardized assessment, 9
Standard scores, 30–34, 332
Standards for Educational and Psychological Testing, 3, 37
Standards for Multicultural Assessment, 234
Stanford Achievement Test, Tenth Edition, 185
Stanford-Binet intelligence scale, 180–182
Stanford–Binet scale, 12
Stanines, 32
State education agencies (SEA), 196
State-Trait Anxiety Inventory, 337
Stereotypic movement disorder, 310
Sternberg's triarchic theory, 173
Stratified sample, 35
Strong Interest Inventory, 216–219
Structured Clinical Interview for DSM-5 (SCID-5), 324
Structured Developmental History (SDH), 277
Structured Interview for the Five-Factor Model of Personality, 240
Structured interviewing, 147–149
Structured personality assessment
  Cannot Say Score (?) Scale, 244
  Conversion Hysteria Scale, 246
  Correction (K) Scale, 245
  depression elevated scores, 246
  Hypochondriasis Scale, 246
  Hypomania Scale, 247
  Infrequency Back [F(B)] Scale, 244
  Infrequency (F) Scale, 244
  Lie (L) Scale, 245
  limitations of, 259–260
  Masculinity-Femininity Scale, 247
  Minnesota Multiphasic Personality Inventory-2, 241–243
  MMPI-2 Restructured Form, 250–252
  Myers-Briggs Type Indicator®, 255–259
  NEO Personality Inventory-3, 252–255
  Paranoia Scale, 247
  Psychasthenia Scale, 247
  psychometric qualities, 249–250
  Psychopathic Deviate Scale, 246–247
  Psychopathology Infrequency [F(p)] Scale, 244–245
  Schizophrenia Scale, 247
  Social Introversion Scale, 248
  Superlative Self-Presentation (S) Scale, 245
  supplementary scales, 248
  Symptom Validity (FBS) Scale, 245
  three-point codes, 248
  True Response Inconsistency (TRIN) Scale, 244
  two-point codes, 248
  validity scales, 244
  Variable Response Inconsistency (VRIN) Scale, 244
Structured personality instruments, 11, 238
Subjective assessment, 10
Substance
  intoxication, 317
  related disorder, 317–318
  withdrawal, 317
Substance abuse
  assessment, 162–164
  motivational interviewing, 163–164
  standardized instruments, 164
Suicide
  instruments for, 157–158
  potential of assessment, 153–158
Suicide Probability Scale, 157–158
Summative evaluation, 333
Superlative Self-Presentation (S) Scale, 245
Supplementary scales, 248
Symptom Checklist-90-Revised (SCL-90-R), 332, 337
Symptom Validity (FBS) Scale, 245
Systematic errors, 39

T

TABE. *See* Test of Adult Basic Education
Tansvestic disorder, 323
Target behavior, 268
TAT. *See* Thematic Apperception Test
Taylor-Johnson Temperament Analysis (T-JTA), 291, 298
Teacher Rating Scale (TRS), 277
Technology-based career assessment, 231–233
Tennessee self-concept scale, 265
Test of Adult Basic Education (TABE), 199
Test-retest reliability, 43–44
Tests
  bias, 93–94
  characteristic curve, 76
  coaching, 211–212
  information function, 54
  instruments *vs.*, 3
  Marital Adjustment, 294
  preparation and performance, 211–212
  sophistication and training, 211
  specifications, 59
Thematic Apperception Test (TAT), 261
Theories
  classical test, 38–39
  domain sampling, 52–53
  generalizability, 52–53
  item response, 53–54, 73–77
  latent-trait, 54
  Sternberg's triarchic, 173
Therapeutic assessment model, 7
Therapeutic Bond Scale, 330
Thought process/content, 165
Three-parameter model, 75–76
Three-point codes, 248
Time sampling, 271
Time Together (TTO), 294
Tourette's disorder, 310
Treatments during counseling, monitoring of, 329–333
Trichotillomania, 314
TRS. *See* Teacher Rating Scale
True Response Inconsistency (TRIN) Scale, 244
T scores, 31–32
Two-parameter model, 75
Two-point codes, 248

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

U

Unidimensionality, 74
United Way, 333
Universal design, 98
Universe score, 53
Unspecified disorder, 305–307
Unstructured interviewing, 147–149
Unsystematic error, 39
U.S. Employment Services (USES), 88
User qualifications, 124
USES. *See* U.S. Employment Services

V

Validity
  assessment instruments, 123
  coefficients, 62–63
  concurrent, 64
  consequences of testing, 70
  construct, 58
  construct-irrelevance, 57
  construct underrepresentation, 57
  content-related, 57–58
  criterion-related, 58
  decision theory, 67–69
  definition, 57
  description, 56–57
  differential item functioning, 62
  expectancy tables, 69
  face, 60
  factor analysis, 61–62
  generalization, 70, 187–188
  instrument content, 59–60
  internal structure, 61–62
  item analysis, 71–77
  multitrait-multimethod matrix, 63
  prediction/instrument-criterion relationships, 63–64
  predictive, 64
  regression, 64–67
  relations to other variables, 62–70
  response processes, 60–61
  scales, 244, 294, 325
  scholastic aptitude tests, 207
  standard error of estimate, 66–67
Values
  Minnesota Importance Questionnaire, 224
  Values Scale, 224–225
Variability, measures of, 25–26
Variable Response Inconsistency (VRIN) Scale, 244
Variance, 26
Verbal assessment, 10
Verbal comprehension subtests, 177
Violence, potential, 159–160
Visual impairments, 111–112
Visual Puzzles, 178
Visual spatial subtests, 177–178
Vocabulary subtest, 177
Vocational aptitude tests
  Armed Services Vocational Aptitude Battery, 208–210
  description, 207–208
  differential aptitude test, 210–211
  WorkKeys, 210
Voyeuristic disorder, 322

W

Wards Cove Packing Company v. Antonio (1989), 88
Watson v. Fort Worth Bank and Trust (1988), 88
Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV), 284
Wechsler Individual Achievement Test-Third Edition (WIAT-III), 200–201
Wechsler Intelligence Scale for Children—Revised (WISC-R), 15–16
Wechsler scales
  fluid reasoning subtests, 178
  processing speed subtests, 179–180
  verbal comprehension subtests, 177
  visual spatial subtests, 177–178
  working memory subtests, 178
Whole-interval time sampling, 272
WIAT-III. *See* Wechsler Individual Achievement Test-Third Edition
Wide Range Achievement Test 4 (WRAT4), 201
Wide Range Interest-Opinion Test-Second Edition (WRIOT-2), 221
WISC-R. *See* Wechsler Intelligence Scale for Children-Revised
WMH-CIDI. *See* World Mental Health-Composite International Diagnostic Interview (WMH-CIDI)
Woodcock-Johnson IV, 183
Working memory subtests, 178
WorkKeys, 210
World Mental Health-Composite International Diagnostic Interview (WMH-CIDI), 324
Worldview, 98
WRAT4. *See* Wide Range Achievement Test 4
WRIOT-2. *See* Wide Range Interest-Opinion Test-Second Edition

Y

Youth Self-Report/11-18 (YSR), 278
YSR. *See* Youth Self-Report/11-18

Z

Z scores, 31

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

Copyright 2017 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s).
Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.